Computers & Society

The Newsletter of the ACM Special Interest Group on Computers and Society Special Issue on 20 Years of ETHICOMP

Special Editor (s): Mark Coeckelbergh, Bernd Stahl, and Catherine Flick

Editor (s): Vaibhav Garg and Dee Weikle

Table of Contents

Special Issue Articles

SIGCAS Officers and Contact Information

Chair Mikey Goldberg SIGCAS E-Mail: chair_sigcas@acm.org	**Vice Chair** Karla Carter SIGCAS E-Mail: vc_sigcas@acm.org
Executive Committee Member-at-large Netiva Caftori E-Mail: netivac@gmail.com	**Past Chair** Andrew Adams E-Mail: aaa@meiji.ac.jp

SIGCAS Computers and Society is an online magazine accessible via the ACM Digital Library. The magazine aims to be an effective communication vehicle between the members of the group. The editors invite contributions of all types of written material (such as articles, working papers, news, interviews, reports, book reviews, bibliographies of relevant literature and letters) on all aspects of computing that have a bearing on society and culture. Submissions may be sent to editors_sigcas@acm.org.

Readers and writers are invited to join and participate actively in this Special Interest Group. Membership is open to all, for US$25 per year, and to students for US$10 per year. The link to join up can be found on our web site, at http://www.sigcas.org .

Copyright Notice

Introduction: A celebration of 20 years of ETHICOMP

Bernd Stahl and Mark Coeckelbergh

In 1995 the first ETHICOMP conference was held in Leicester, England, organised by Terry Bynum and Simon Rogerson. Its purpose was to provide a forum to discuss ethical issues around computers. Twenty years later we met again in Leicester to continue this conversation. The changes in information and communication technology (ICT) during these 20 years have been dramatic. While computers used to be bulky and easily identifiable machines, we now have small smart devices, the internet quickly developed and has changed significantly, and ICT now pervades all walks of life, from the way we work and communicate to study, undertake childcare and choose partners. As a consequence many of the concerns of 1995 have deepened and many new ones have arisen.

During ETHICOMP 2015, we discussed ethical and social issues raised by contemporary computing and look at ways of identifying and addressing them in the future. The conference aimed to be practically relevant and bring together the various communities involved in the development, implementation, use of computing and reflection on it in its various guises. The conference is based on the belief that the ETHICOMP community, together with other associations and groups, need to work together to enable the benefits of computing to prevail, while rendering its downsides and ethical ambiguities visible and more subject to public debate than is the case today.

To structure the discussion we had the following tracks: Researchers' issues in Computer Ethics and Information Ethics studies, Social Impacts of Snowden's Revelations: Worldwide Cross-cultural Analyses, Responsible Research & Innovation in Industry, ICT and Society: Social Accountability, Professional Ethics and the Challenges of Virtuality and the Cloud, Teaching and professional ethics, Robot Ethics, an Open track (topics of relevance that did not fit any of the themes), and Novel formats such as film.

We received an impressive response to our call for papers and could offer conference participants a broad choice of interesting topics. Issues discussed at the conference included machine learning approaches to moral judgment, computer ethics as empirical ethics, the impact of Snowden's revelations in countries such as China and Japan, research and responsible innovation in the area of brain-computer interfaces, nudging and waste management, facts and fiction in the cloud, the history around ICT and society, how to prepare information systems students to meet global challenges, the ethics of human-chicken relationships, the ethics of smart transport, cryptocurrencies as narrative technologies, augmented reality, trust and mobile government, the invisible robots of global finance, the human in the ethics of robotics, robots, ethics, and language, and robots and trust.

The present conference proceedings offer the reader an overview of most of the papers presented at the conference. We hope that they will contribute to many more years of good discussions about computer ethics and related fields of research.

Most previous ETHICOMP conferences had paper proceedings. These weighty tomes were appreciated by many authors but they were not very visible and easy to access. They also created additional costs. We therefore decided to explore different ways of disseminating the good work presented at ETHICOMP and are grateful for the support by Andrew Adams and the ACM SIGCAS newsletter editors for allowing us to use this outlet to make the papers much more widely available than they were in the past.

This conference, like any event, is the result of contributions of many individuals. We would like to use this opportunity to thank them. Special thanks is due to the members of the ETHICOMP Steering Committee, the Track Chairs and members of the Programme Committee. We would also like to thank the individuals who contributed to the organisational efforts, including Gurminder Badan, Liz Stokes, Christian Hansen, Tilimbe Jiya, Michelle Brown, Kathleen Richardson, Mamadou Bamba, Shamimaa Ali and Paul Keene. And of course we would like to thank the authors and presenters who made the conference possible and provided the content for this special issue.

Generation Process of Gaze by the Surveillance Camera: Case of kamagasaki of Japan

Hiroshi Koga
Kansai University
2-1, Ryozenjicho, Takatsuki
Osaka, Japan
Telephone number, incl. country code
koga@res.kutc.kansai-u.ac.jp

ABSTRACT
In this paper, the author considers the process of privacy issues in the surveillance society has been formed on the case study of *Kamagasaki* area in Osaka, Japan. The author has adopted a viewpoint of social constructionism. Kanagasaki district is famous for the area that saveillance cameras were collectively installed in a specific area for the first time in Japan. From a perspective of "sociology of social problem", we discuss the transition of view of CCTV (Closed Circuit Television) or surveillance society.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues

General Terms
Human Factors

Keywords
CCTV, surveillance, Social Construction of Technology

1. INTRODUCTION
This paper aims to discuss transition of view of CCTV (Closed Circuit Television) or surveillance society in *Kanagasaki* district, where surveillance cameras were collectively installed in a specific area for the first time in Japan, from a perspective of social constructionism. In particular, "sociology of social problem"[1] is adopted [1], [2].

It was in 1966 when CCTVs were installed in the area for the first time[2]. According to Osaka prefectural police, there were installed intended to prevent crimes on the streets in the district. Subsequently, 15 units in total were additionally installed by 1983. Moreover, Osaka prefectural police announced in 2014 it would newly install 32 units while replacing existing 13 cameras by high performance ones [3], [4]. Some attorneys insist that more than 100 units have been installed in reality. Such a place that CCTVs have been installed collectively to that extent is rare in Japan.

Besides, constructionism in social problem study is a way of thinking "to conceptualize social problems not as a specific objective state but as a string of people's defining activities over social problems and to discuss processes of such activities as well as methodology to organize them" [5]. In this paper, I have adopted a viewpoint of social constructionism which introduced the following insights of methodology; 1) not to use conception as a tool for explanation, 2) to think not based on cause and effect but by successive and sequential perspective, and 3) activities, context, and those who lead such activities to be configured reflexively (see [5]).

2. BACKGROUND
Recently, "sensory anxiety", that reassurance of the society has been lost, has been insisted to increase in Japan [6]. In order to relieve sensory anxiety, it is required to revise customs and human relationship within regional communities. However, it is a strong trend to set up CCTV in public space (shopping area, parking lot, and on the streets) as a tool to easily relieve sensory anxiety.

[1] Social constructionism is the positon of sociology that "reality (reality of social phenomena, facts and realities that exist in society, meaning) are those that have been created in people's head (in the emotions and consciousness), it does not exist it away it.

[2] Case study in this paper was described based on the books and published archives. In particular, we refer to the following literature. Kamagasaki Siryo Center. 1993. *Kamagasaki: History and Current*, Sani-chi Syobo (in Japanese); Honda, T. 2006. *Kamagsaki and Gospel*, Iwanami-Syoten (in Japanese); Haraguchi, T. *et al.* 2011. *Kamagasaki no Susume*, Rakuhoku. (in Japanese) .

Those who have a point of view to tolerate or promote CCTV have pointed out that it is useful for crime investigation and capable to protect crime victims (find out criminals) with an advantage of expected crime deterrent effect [7]. They also insist it should be allowed to expand police authority power in order to secure safety and security.

In a surveillance society theory, ambiguity of digital surveillance is pointed out. In this age when surveillance objects have been subdivided from the human body and shifted to digitalization, surveillance has been enhanced not only in control but also care aspects (ambiguity of surveillance)[8]. For example, in case of shopping, data surveillance is understood to have a function to regulate actions based on an assumption that people keeping a point card (loyalty card) are bonded and/or retained by a firm that issued the card. Therefore, surveillance is understood as "a society depending processes of governance and control on ICT" in "a state that non-physical surveillance has penetrated into the society" [9]. However, there are not a few critical opinions for the advent of surveillance society, in particular, installation of CCTV in public space [10], [11].

One of criticisms against CCTV surveillance in public space is suggested from a point of view that CCTV may violate privacy. It was by Tokyo District Court ruling in 1964 when a right to seek damages was recognized for violation of the right to privacy for the first time in the legal story of Japan. It was so called "Utage-no-ato (After a Feast)" event. "A right that private life shall not be disclosed without good reasons" was admitted as a sort of personal right against a model novel written by Yukio Mishima and the court ruled the novel had violated the right and ordered to pay damages (such frame of mind has been succeeded subsequently in trials such as "Ishi-ni-oyogu sakana (Fish swimming through stones) event by Mi-Li, Ryu).

CCTV installed in a public space, however, is different from peeping. Originally, traditional definition of privacy was a right to be let alone just in a private space and public space was not supposed. Therefore, such concept was suggested that people have a right to control their own information disclosed in a public space.[3]

Another criticism insists that surveillance activity may impair a balance between *symbolic* and *die Diabolik* (Satanship). N. Luhmann contrasted synmbolish oriented to integration, harmonization, order, and safety with diabolish with functions to create differences to be separated and to destroy integration and order [12]. By employing this concept, the act of surveillance is believed to undermine synbolish attitude and accelerate diabolish attitude. Dependence of security and safety on surveillance is believed to make "social capital" and "active trust" vulnerable[4].

[3] For example, see Munesue, T. 2003. Surveillance Cameras and Privacy, *ITE Journal*. 57(9), 1076-1077. (in Japanese); Sinpo, F.: The definition of the term 'life-log' and related legal accountabilities: The appropriateness of using personal records for commercial purposes, *Johokanri*, 53(6), 295-310, doi: 10.1241/johokanri.53.295 (in Japanese); O'Hara K, *et al.* 2006. Memories for life: a review of the science and technology, *J. Royal Soci. Interface*, 3, 351-365. (in Japanese)

[4] Refer to the discussion of "social capital" for such a challenge. For example, see Putnam, R. D. 1995. Bowling Alone: America's Declining Social Capital, *Journal of Democracy*. 6. 65-78.

The former criticism is a critical attitude against surveillance activity itself. However, it has been pointed out that such stance contains both criticisms against values achieved by surveillance and surveillance itself in a mixed manner.

Studies have been strenuously made for the latter criticism seeking for solutions (mainly in a research field of image processing) by establishing a system not to invoke self-information control right[5]. In such studies, based on a concept that whether self-information control right is violated or not depends on contexts and methods for utilizing image data recorded by CCTV, it is believed to be able to prevent privacy problems by 1) introducing a full-automatic system with no personal identification and not mediated by human, and 2) anonymizing image data obtained (for example, by replacing human by a bar-like mark).

As roughly reviewed above, various discussions have been developed for surveillance in public spaces by CCTV including concept definition of digital surveillance and privacy as well as technological prevention measures. In this paper, the author would like to clarify how a social problem appeared on "surveillance" existing on a background of such discussions. Then, I hope to be able to provide some clues for discussing IT ethics on CCTV by sequentially describing the process that social problem of surveillance was established.

3. CASE OF KAMAGASAKI IN JAPAN

3.1 Short History of Kamagasaki

As above mentioned, so-called "*Kamagasaki*" district of Nishinari-ku, Osaka is the place where surveillance cameras were collectively installed in a specific area for the first time in Japan. In order to explain why CCTVs were collectively installed in the district and the background, we roughly review the history of the district in advance.

Kamagasaki district is generally recognized to be an area around south of JR Shin-Imamiya station (Hagino-chaya, Taishi, Hanazono), Northeast part of Nishinari-ku, Osaka. The area is merely 0.62 km^2.

However, place-name of *Kamagasaki* is not on a map, in other words, *Kamagasaki* is an "uncharted town." Place-names including *Kamagasaki* have disappeared from maps by revision of place-names in 1922. Nevertheless, the place name of *Kamagasaki* remained. The reason is deeply related with a city planning of the district. According to a modern city planning in early 1900, cheap lodging houses for day employees scattered in other areas were moved to *Kamagasaki*. Such accommodation facility is called as *Doya* in Japanese. The word *Doya* created by reversing *Yado* (lodging house) is casually used still today.

From a burned-out site considerably damaged by the World War II, the district restored in 1950's to an extent that number of cheap lodging houses exceeded that of before the war. Due to poor environment of the day employees, *Kamagasaki* was regarded as a slum area. Subsequently in 1961, Osaka city

[5] A large number of engineering studies have been reported. For example, please reffer Sekiguti, T and Kato, H. 2006. Proposal and Evaluation of Video-based Privacy Assuring System Based on the Relationship between Observers and Subjects, *Journal of Information Processing*, 47(8), 2660 – 2668. (in Japanese)

announced to implement supports of life, learning, and employment as a poverty program for *Kamagasaki*.

However, complaints of most day employees were made to police and placement system rather than to poverty. There is a well-known episode of their complaints against police. That is a story that when they ran into a police station saying, "We have been robbed by a mugger," police officers made them irrelevant responding, "it's your own fault if you were sleeping on the street." The employment agency at *Kamagasaki* is known as the only agency that does not offer jobs in Japan. Those who offer jobs to day employees are labor sharks. But they often cause problems such as soliciting kickbacks and non-payment of wages. Complaints caused by these reasons were accumulated between day employees.

In 1961 amid such situation, a "riot" broke out in *Kamagasaki*. It was induced by a way of handling of a traffic accident in which a day employee was injured. The injured person died because police officers heading for the accident site had neglected urgent handling. With destruction of police boxes within the district, the riot provoked by day employees continued for a few days including stone throwing toward *Kamagasaki* police station. Subsequently, similar riots happened two times by 1963. The author believes that one of the backgrounds of the riots lies in failure of recognition by the government regarding problems that day employees are facing with.

Taking the opportunity of the riots, the government has recognized the day employees as a problem. It became to deal with day employees themselves as a problem not to resolve a problem (poverty) between them in the past. In exchange for disappeared slum area measures, maintenance of security became the main purpose of *Kamagasaki* measures. In addition to transfer of family households outside the district, dockworkers that were thought to be with fiery temperaments were moved. The police responded focusing on overseeing day employees. Aiming at establishing a system to positively abort crimes, Osaka Prefectural Police increased 20 officers to apply a 421-officer system and installed four CCTVs on the streets. The objection voiced by the day employees took a vicious form of riot resulted in diminishing their own status to "a person who disturb security, i.e. object of surveillance."

In the wake of fifth riot, the national and local governments and news media agreed to call *Kamagasaki* district as "Airin (meaning of lovable neighbors in Japanese) district" on May 1966. Day employees who live in the district, however, use the name of "*Kamagasaki*" still now and seem to avoid using "Airin" as a derogatory term imposed from governments. In contrast, mass media often use "Airin." In addition, not a few residents appreciate the term as a name to sweep the image of riot place. In some cases, it is even said that a crack may be generated by which name to be used. In this article, we intentionally adopt the name of *Kamagasaki* in convenience for developing discussion.

Besides, it is said as one of reasons for adoption of new name, Airin district, that there was also a purpose to alleviate the image of riot in order not to cause any problem for recruiting day employees for construction of Osaka Expo venues. With increased small-sized cheap lodging houses for one person, *Kamagasaki* has become a town of one-person male workers at last.

3.2 Problem Formation Process in Kamagasaki CCTV Lawsuit

Those who hold a trouble of riot (governments and police) tried to prevent riots and settle a security problem by leaving just single non-dock workers in *Kamagasaki* (separation of workers) and installing CCTVs. In particular, 13 times of riots broke out just for three years in 1970's. Based on a suspicion of agitation by leftist activists, such speculation ran rife that CCTV was aimed at monitoring activists. Response of day employees to such speculation was indifferent or tolerant. Problem for them was dissatisfaction cast toward the placement system and police. CCTV was nothing but a source of stimulation for dissatisfaction. On the other hand, organizations working on activities to support day employees on a daily basis, different from activists to stir riots, also took a dim view of surveillance by CCTV. The problem was recognized when a CCTV was installed so as to film at an entrance of a base building of an organization. Even though the police suggested that installation of CCTV was just for crime prevention, disbelief in the police was amplified partly because CCTV was not installed near gang offices in *Kamagasaki* district. It is not difficult to imagine such a suspicion that "it may be a surveillance of human right activities" have appeared in such situation. Such discontent as a feeling that their every action is kept under surveillance was brought into an arena called as court. The problem was handed over from an influential supporting organization called as a joint labor union to an attorney through their referral network. By the attorneys, the problem translated into a text persuasive enough to discuss in the courtroom. "Shift of arena", which is often described in social constructionism, was performed [3]. "Surveillance of supporting activities for labor issues and human rights" was replaced by a problem of "privacy violation" (redefinition of problem). Thus, a controversy erupted over application of CCTV whether it works for security maintenance and crime prevention or violates human rights. Amid the dispute, "a lawsuit demanding camera removal and compensation money" was filed by a plaintiff of 12 persons including a chairman of joint labor union and the district against Osaka Prefectural Police as the defendant.

The arena in the name of court forced the police to complement their argument. They insisted that installation of CCTV had some effect for maintenance of security and crime prevention. Further, they repeatedly insisted that CCTV was not for investigation of cases but for crime prevention saying, "We don't record the places and so, it is just the same as patrol on the public roads."

The court ordered the police to remove the CCTV installed in front of the building the supporting organization resided. And it has suggested a judgment that usage of CCTV installed on a public road as a part of intelligence activity is basically within the discretion of the police. Further, the following items were indicated as conditions for installation and use of CCTV. That is, 1) legitimacy of purpose, 2) objective and specific necessity, 3) adequacy of installing situation, 4) expected effect by the installment and use, and 5) reasonability of method of using should be considered. In addition, the ruling also indicates that judgement of legality of CCTV shall be considered individually, resulting in removal of only one unit as described above among 15 units. (Osaka District Court, April 27th, 1994, judge, 1515, P. 116 / The ruling was kept by both the High Court and Supreme Court.)

The ruling of Osaka District Court took attention as a breakthrough case in aspects that individual judgement for a group of cameras was emphasized and that criteria of illegality were suggested.

A problem established in an arena in the name of court was resolved in a form of ruling. However, the ruling was understood as different solutions between plaintiff and defendant sides. The plaintiff side took it as a winning lawsuit even though removal of just one unit was unsatisfactory. The defendant side, i.e. the police had a sense of victory even allowing for the order of removal for one unit because the remaining CCTVs were legally adequate (based on an assumption of no recording) and criteria of CCTV installation were suggested. In fact, it became to be cited as a legal foundation when introducing CCTV mainly to down towns subsequently. Different interpretations were generated for a problem and its solutions.

Meanwhile, no riot erupted during the lawsuit and passionately prosperous 80's; however, 23rd riot broke out in 1990 induced by a scandal of the police. The latest 24th riot broke out in 2008, but no riot has erupted since then.

It may be deeply related with aging of workers living in the district. Due to advanced aging in day employees recruited from across Japan in 1970's, they have been trapped in a difficult situation to get jobs. Workers who lost income were forced out to live on the streets. The government regarded such street people as a cause of security deterioration. In short, a new problem has been recognized in *Kamagasaki*.

Further unfortunately, persons of goodwill appeared who brought their disused articles in to *Kamagasaki* aiming at supporting street people. However, their goodwill was nothing but self-righteousness supported by an image that "what was just a waste to him/her might be helpful for them." The government was bewildered by eruption of new problem of illegal dump of wastes. Goodwill from outside the area resulted in deterioration of security. With a concept that sweeping street people may resolve illegal dump of wastes, the government made efforts to sweep street people. Some of them were forcibly expelled.

Under such situation, the supporting organizations have urged them to become independent by fixing their addresses to receive welfare. Some cheap lodging house business operators renovated their facilities into apartments to provide welfare recipients with residences. It is so called "welfare apartment." Among former day employees who receive welfare being unable to get jobs due to aging, however, some of them have become to drink alcohol even during daytime. Morning for day employees engaged in physical work starts earlier. Therefore, none of them drank too much at night. But, they say they have become to drink during daytime because they have nothing to do during daylight after retirement. On top of that, eateries to serve alcohol for such elderly have increased. Once in *Kamagasaki*, workers left in the district in the daytime except for rainy days were "unemployed people." Therefore, there was no eatery to serve alcohol during daytime.

In *Kamagasaki* as of now, chain of new problems have been generated such as street people, illegal dump of wastes including cases caused by goodwill, increase in welfare recipients (and its accompanying financial problem of the government), and increase in drunken persons. In addition, drug trading by gang groups has been accused of and it has further worsened the image of dark side for Kamagadaki. The government has embarked on a cleanup campaign to eliminate dark side within Osaka city. Along with increased routine patrols by the police, removal of tents and illegally parked two-wheeled street carts of street people has been performed thoroughly. Full-fledged cleanup activity has been under way for illegal dump. A part of street people received compensation by participating in the cleanup activity. Setup of new CCTV is understood as a part of the activity. The project has been announced in 2014.

The government may have understood that it is possible to expand installation of CCTV as long as the criteria suggested by the previous ruling are complied with. According to the argument of the government, objects of surveillance are neither the supporting organizations nor workers but potential criminals.

Nevertheless, supporting organizations and attorney group felt discomfort for additional installation of CCTV. As countermeasures against homeless people have been positively taken by the government along with increased homeless workers from lack of budget for living expenses affected by the long-lasted downturn since 2000's, a suspicion has erupted whether CCTV is used for surveillance of homeless people. As if to support it, the government made every effort to create an environment difficult to live on the streets by taking measures for thwarting life on the streets such as installation of large flowerbeds on the sidewalks or sprinkler on building walls to flood the streets, and lockout of parks. CCTV was understood as a part of it.

It has been reported on May 2015 that organizations and attorneys supporting *Kamagasaki* would take judicial procedures demanding CCTV removal. When a "problem" is transferred to an arena of court, "conflict over discrimination for homeless people" is required to be re-interpreted. In such a case, the threat expected to be emphasized by reflecting the recent CCTV controversy. For example, in 2014 at the same Osaka city, a verification test of CCTV planned for a commercial building integral to JR Osaka station was temporarily suspended. Outlook of the test was as follows. By automatically identifying parts corresponding to faces of passersby from images taken by using 90 units of CCTV, the test was intended to make use of the data for marketing and planning of evacuation routes by recognize how long they stayed in the building and facilities and how they moved. With surging concern for privacy violation, however, the test was postponed and started in 2015 by significantly reducing the scale.

It has been also indicated that CCTV has less effect for reducing crime itself. Even though it seems to be contradictory at a glance, it has been pointed out that surveillance may have an effect to shrunken activities of those who are monitored. The assertion of the police that CCTV without recording is the same as the routine patrol is lack in persuasiveness in a technological environment where individual recognition is possible.

Moreover, in consideration of activities such as installation of various barriers, collection and disposal of tools of homeless people (such as street two-wheeled cars for their living articles, or collection of used paper and cans) in a name of cleanup activity in a way not to support homeless people by welfare policies but to physically thwart life on the streets, it is believed such activities may have created a regional sentiment that homeless people should be abhorred. It is understood as a stance to create an atmosphere of mutual surveillance by residents. In consideration of these contexts, it is hard to avoid thinking that the government desires to regard CCTV as "a tool to monitor homeless people" or "a symbol of attitude expression to expel them and their support activists." However, it is difficult to directly appeal removal of

CCTV with political nature in the court. Therefore, it is expected that another dispute over privacy protection would erupt.

4. DISCUSSION

It is not too much to say that controversy over expansion of surveillance camera installation has a political nature just like an extremely low bridge thrown over the park road in Long Island. It cannot deny a possibility that surveillance camera as a device for representing a sense of discrimination may create synopticon or perioptic from panopticon surveillance. If such a sentiment was embedded in the society that "you should not become a person like that", it cannot deny a risk that a society in which people monitor each other may be created apart from surveillance camera. It is a paradoxical phenomenon to produce perioptic surveillance from panopticon surveillance. Perspective of social materiality may be useful to explain these paradoxical phenomena. I would like to demonstrate its availability as a challenge in the future through further detailed analysis.

5. ACKNOWLEDGMENTS

This study was supported by JSPS KAKENHI Grant Numbers 26380550, and by Kansai University's Domestic-Research-Program (April-September/2014) and Kansai University's Overseas-Research-Program (April-September /2015).

6. REFERENCES

[1] Kitsuse, J. I. and Spector, M. 1987. *Constructing Social Problems*, Aldine de Gruyter.

[2] Winner, L. 1986. *The Whale and the Reactors*, University of Chicago Press.

[3] Mainichi Sinbun (newspaper). 13 January 30, 2014. Osaka Nishinari: Street Concerning Measures Against Crime, in five years ¥ 500 million to investment, (in Japanese) It was able to browse the web, but now it cannot access (http://mainichi.jp/select/news/20140131k0000m040133000c.html).

[4] Osaka prefecture. 2015. Governor of fear management policy description (Summary), (in Japanese), It was able to browse the web site. http://www.pref.osaka.lg.jp/kikaku/hatsugen/250221.html

[5] Nakagawa, T. 2003. Translation" and the Mosaic of Legal Reality: A Few Suggestions according to a Claims-Making Approach, The Sociology of Law, 2003,58, 79-97,273.

[6] Kawai, M. 2004. The paradox of collapse of safety myth: law sociology of security, Iwanami Syoten (in Japanese).

[7] Esita, M. 2005. New dimensions on surveillance society, Jounal of Policy Studies, 20 (July 2005), 206-207. (in Japanese)

[8] Deleuze, G. 1990. *Pourparlers 1972 – 1990*, Minuit.

[9] Lyon, D. 2003. *Surveillance as social sorting: Privacy, risk, and digital discrimination*. Psychology Press.

[10] Ogura,T. (2006), Electronic government and surveillance-oriented society. In David Lyon, Theorizing Surveillance, Chapter 13.

[11] Abe, K. (2004), Everyday Policing in Japan: Surveillance, Media, Government and Public Opinion, International Sociology, 15, pp.215-231.

[12] Luhmann, N. 1968. *Vertrauen: ein Mechanismus der reduktion sozialer Komplaxitat*. Ferdinand Enke Verlag.

Perceptions of incompetence in the ICT workplace

Yeslam Al-Saggaf
School of Computing and Mathematics, Charles Sturt University
Boorooma Street, Wagga Wagga
NSW 2678, Australia
+61 2 69332593
yalsaggaf@csu.edu.au

Oliver K. Burmeister
School of Computing and Mathematics, Charles Sturt University
Panorama Avenue, Bathurst
NSW 2795, Australia
+61263386233
oburmeister@csu.edu.au

John Weckert
Centre for Applied Philosophy and Public Ethics, Charles Sturt University
Brisbane Avenue
Barton, ACT, 2600, Australia
+61 6272 6284
jweckert@csu.edu.au

ABSTRACT
The aim of this study is to examine incompetence in the Australian ICT workplace from the perspective of Australian ICT professionals. The data collection for this project included conducting a quantitative survey, conducting qualitative interviews and conducting focus group discussions with key informants. Of the 2,315 respondents who participated in the survey, the MRF analysis revealed that incompetence was ranked fifth from the top of a list of the 57 most common ethical problems experienced by ICT professionals (N=750, 35.9%). An inspection of the results of the cross tabulations revealed that 34.8% described their occupational category as manager and 29.1% indicated they were consultants. The GLM has found a significant relationship between the choice of incompetence and occupation (Deviance = 23.15, Df = 6, P=0.0007) suggesting occupation, among other things, does predict the choice of incompetence. The findings from the qualitative interviews are consistent with the above findings. A cross referencing of the interviewees responses that addressed the issue of incompetence during the interviews against their occupation has revealed that consultants had more to say on the topic than any other occupation (20.8% or 10 of the 48 references). This is followed by managers who accounted for 14.6% (7 of the 48 references). These findings indicate that the experienced professionals have a greater concern about incompetence than others; an observation that the findings from the focus group interviews have also confirmed. Obtaining such findings would not have been possible had only one method been used.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues

General Terms
Human Factors

Keywords
Professional ethics, generalised linear models, deviance, social accountability.

1. INTRODUCTION
This study examines incompetence in the Australian ICT workplace from the perspective of Australian ICT professionals.

This study is part of a large research project that looks at professional ethics in the information and communications technology (ICT) workplace in Australia.

This research project employed a mixed methods approach to find out Australian ICT professionals' perceptions regarding what helps them identify ethical problems in the workplace and solve them. The main research question was: What are Australian ICT professionals' perceptions regarding the ethical problems they face in the workplace, and how are these problems resolved? The project involved three phases. The first phase involved a quantitative survey of members of the Australian Computer Society (ACS), administered using SurveyMonkey.com. The second phase of the study involved a set of semi-structured in-depth interviews with 43 participants selected from those who responded to the first phase. The third and final phase of the research involved focus groups with senior ICT professionals, to determine the composition of a web resource, to better match strategies for solving ethical problems to those problems, and to refine those strategies to be most effective.

This paper reports on incompetence using data collected from all phases of the project. The study found that incompetence was ranked fifth from the top of the list of the most common ethical problems experienced by ICT professionals. The pervasiveness of ICT in society, through mobile technology, virtuality, cloud computing and more, demands that this industry addresses how to gain and maintain public trust. Therefore it needs to recognize and address the problem of incompetence. The next sections looks at the notion of incompetence from the literature and then moves to detail the methods and findings of this study.

2. THE NOTION OF INCOMPETENCE
A search within the recent academic literature has revealed that with the exception of the two studies discussed below, there are hardly any studies on incompetence from the perspective of ICT professionals. Similarly, a review of several recent ICT ethics books has revealed that while competence, not incompetence, was mentioned once or twice in passing, incompetence as an ethical issue has not been covered in these books. In addition, the search of journal databases has revealed that very few studies looked at incompetence in other disciplines. This suggests that there is a gap in the literature relating to incompetence in general and from the perspective of ICT professionals specifically. This article seeks to address this imbalance in the literature by focusing on incompetence in the Australian ICT workplace.

The few studies that the search of the literature has located have shown that there is no agreed upon definition for incompetence. It appears that the term means different things to different people

and is often influenced by the discipline where the term is used. Within the construction industry, incompetence is "the inability of a contractor to deliver goods and services contracted for to the specifications required by a client" [1]. Thus if the delivered goods and services don't match the requirements specified in the contract, the contractor can be labeled incompetent. In the nursing sector, incompetence is not an honest mistake by a nurse or an isolated event; rather a pattern of behavior [2]. Given giving the wrong medication or the incorrect dose can result in the death of a patient, the number of times a nurse makes a mediation error influences her peers' decision to report her for wrongdoing [2].

Within the context of system development Avizienis et al. [3] view incompetence as a potential threat to the dependability of a system, which is defined as "the ability to deliver service that can justifiably be trusted". Specifically, incompetence is considered one of the causes for non-malicious human made faults, which are introduced without a malicious objective. Avizienis et al. [3] argue that harmful mistakes and bad decisions can be made by persons who lack professional competence to do the job they have been asked to do. But Avizienis et al. [3] argue that incompetence is not limited to individuals; an entire organization can be considered incompetent if it did not have the organizational competence to do the job. In an Australian context an example of organizational incompetence is the faulty Health Payroll system that IBM delivered to the Queensland Government in 2010 [4].

In the context of information technology management Baba [5] views competence, which encompasses both technical skill and interpersonal skill, as an important dimension of trust. According to Baba [5]

> If we trust in the competence of another, we expect that he or she has the requisite knowledge, skill, and personal characteristics (e.g., dependability) needed to perform an action in a way that results in a positive outcome for us (e.g., we trust our surgeon to perform the operation in a competent manner).

From the above, it is clear that competence is a prerequisite for trust. Baba [5] notes that the reason for the general public distrust of information technology is the public perception that those in positions of authority are incompetent. Competence here is couched in terms of the requisite knowledge, skill, and personal characteristics (e.g., dependability).What is also interesting to note from this account is that dependability here is that of a person; not of a system, as Avizienis et al. [3] above had indicated. But the research that Baba [5] conducted has found that distrust based on perception of incompetence is one of the main reasons for users' objection to new information technology. This is in line with what Euchner [6] has found. Euchner [6] found that one of the factors behind the underlying resistance to innovation is "skilled incompetence". It seems that distrust on the basis of perception of incompetence of ICT people leads to distrusting the technology itself.

It would appear from the above views about incompetence that trust and the dependability of a system or a person are common denominators. However, do the two refer to the same thing? Merriam Webster Dictionary [7] definition of trust as the "assured reliance on the character, ability, strength, or truth of someone or something" or "one in which confidence is placed" suggests that dependability is one of the elements of trust. Good [8] and Rempel [9] consider dependability a condition that must be met for an individual to trust another. Weckert [10] agrees with this view. In his philosophical account on trust, Weckert [10] argues that reliance and being reliable are two of many meanings associated with trust but while reliance is part of trust the two are different. Weckert argues "I rely on my glasses to see, but in no interesting sense do I trust them". The responses of the interviewees in this study will be carefully checked for the presence of the trust and dependability themes. It should be noted that in the ACS Code of Professional Conduct[1], competence is listed as one of the six core ethical values. According to the Code

> Accept only such work as you believe you are competent to perform, and do not hesitate to obtain additional expertise from appropriately qualified individuals where advisable. You should always be aware of your own limitations and not knowingly imply that you have competence you do not possess

In light of the scarcity of research on incompetence in the literature especially from the point of view of ICT professionals, this article will take as incompetence what the respondents in this study perceive as incompetence.

3. METHODS AND FINDINGS

3.1 The quantitative survey

3.1.1 Survey procedure

Phase 1 involved a survey questionnaire administered using SurveyMonkey.com, to allow the participants to fill in the questionnaire and return it over the internet.

All active ACS members (approximately 18,600) were invited to participate in the web-based survey by direct email sent to them by the ACS once on 12 September 2013. The survey was closed on 6 November 2013 after the response rate reached 12.4%. The online questionnaire was prefaced by the ethics consent sheet (including assurances of anonymity) and a description of the study. The questions comprised both closed ended and open ended questions. This study reports only on the closed ended questions.

3.1.2 Sample

A total of 2,315 participants completed the questionnaire. Of the 2,315 respondents who participated in the study 84.5% (N=1940) were males, and 15.5% (N=356) were females. By age, 11% (N= 254) of the respondents indicated that they were between 31 and 35 years; 12.4% (N=287) indicated that their age fell between 41 and 45 years; 14.1% (N=325) said their age fell between 51 and 55 years; and 12.2% (N= 282) indicated that they were 61 years and above. In terms of the respondents' geographic location where they work most of the time, 10% (N=247) selected ACT; 30.4% (N=696) selected NSW; 12.2% (N=279) selected Qld, 5.5% (N=120) selected SA; 25.4% (N=581) selected Vic; and 9.5% (N=218) selected WA. According to the survey results, of the 2,315 respondents who participated in Phase 1, 33.8% (N=698) described their occupational category as manager, 14.8% (N=307) said they were developers; 24.3% (N=502) indicated they were consultants and 13.3% (N=277) said they worked in technical support. The results have also shown that 90.6% (N=2069) work in a capital city or metropolitan area while the 9.4% (N=215) live in regional areas. In terms of industry type 12.3% (N= 244) work in Education, 11.9% (N=235) work in Finance, 15.4% (N=304)

[1] http://www.acs.org.au/__data/assets/pdf_file/0014/4901/Code-of-Professional-Conduct_v2.1.pdf

work in Government, and 34.2% (N=676) work in ICT. The majority of respondents work in the private sector 57.4% (N=1,282) with only 27.5% (N=614) working in the public sector. The majority of respondents are permanent full-time employees 65.4% (N=1,388) and only a few of them are either independent consultants 5.2% (N= 110) or self-employed 5.3% (N=112). The average number of years of experience for all respondents is 19 years. Finally, in terms of qualification, 36.4% (N=794) have a bachelor degree and 34.6% (N=756) have higher degrees.

3.1.3 Statistical analysis

3.1.3.1 Multiple Response Frequency (MRF) analysis and cross tabulations

The first question this analysis tried to answer was: what is the ranking of incompetence in terms of the most frequently faced problem of the 57 problems the participants were asked about? Since the question about the most common ethical problems allowed respondents to select more than one answer, a Multiple Response Frequency (MRF) analysis was judged to be the most appropriate analysis technique. In addition, cross tabulations were also performed to see if there are differences in responses based on the demographic information. The findings from the MRF analysis and the cross tabulations are summarized below.

Of the 57 ethical problems listed for respondents to select, the MRF analysis revealed that incompetence was ranked fifth from the top of the list of the most common ethical problems experienced by ICT professionals (N=750, 35.9%); with "Compromising quality to meet deadlines" being the highest on this list. For information on the top 10 list see [11]. An inspection of the results of the cross tabulations revealed that 85.9% of the ICT professionals who selected incompetence as the most common problem are male with the remaining 14.1% being female. By age, 9.2% indicated that they were between 31 and 35 years; 13.8% indicated that their age fell between 41 and 45 years; 15.4% said their age fell between 51 and 55 years; and 14.6% indicated that they were 61 years and above. In terms of geographic location where they work most of the time, 11.9% selected ACT; 29.3% selected NSW; 14.9% selected Qld, 5.4% selected SA; 23.6% selected Vic; and 9.6% selected WA. According to the survey results, 34.8% described their occupational category as manager, 14.1% said they were developers; 29.1% indicated they were consultants and 10.9% said they worked in technical support. The results have also shown that 92% work in a capital city or metropolitan area while the 8% live in regional areas. In terms of industry type 11.9% work in Education, 11.8% work in Finance, 17.6% work in Government, and 33.5% work in ICT. The majority of respondents who selected this problem work in the private sector 58.8% with only 29% work in the public sector. The majority of respondents are permanent full-time employees 66.9% and only a few of them are either independent consultants 6.9% or self-employed 5.9%. The majority of those who selected incompetence as the most common problem have 30 years of experience (9.5%). Finally, in terms of qualification, 34.9% have a bachelor degree and 35.4% have higher degrees. From the above a comparison can be made between a typical respondent to the survey questions and a typical selector of incompetence as the most common problem. See table 1 below for this comparison.

Table 1. A comparison between a typical respondent to the survey and a typical selector of incompetence as the most common problem

Demographic information	Typical respondent to the survey questions	Typical selector of incompetence
Sex	Male	Male
Age	51-55	51-55
Years of experience	19	30
State/territory	NSW	NSW
Geographic location	Capital city or metropolitan area	Capital city or metropolitan area
Self-described category	Manager	Manager
industry	ICT	ICT
sector	Private	Private
Employment type	Permanent full-time	Permanent full-time
Highest qualification	Degree	Higher degree

3.1.3.2 Generalized linear models

Given a typical selector of incompetence differed from a typical respondent to the survey questions in terms of years of experience and qualification, we predict that years of experience and qualification will predict the choice of incompetence. The responses to the choice of incompetence is binomial (recorded as a Yes/No) whereas all the demographic variables are categorical. For this reason generalised linear models (GLMs) were fitted to test our predictions and investigate other relationships between these predictor variables and the binomial response. The GLMs were carried out on the data using R (version 3.0.2 (2013-09-25)) open-source statistical software. It is a requirement of this analysis that there is no evidence of overdispersion in this model. In all cases this requirement has been verified.

The analysis of deviance has revealed that there is a significant relationship between the choice of incompetence and age (Deviance = 40.55, Df = 8, p=0.0001), state (Deviance = 14.25, Df = 8, p=0.076), occupation (Deviance = 23.15, Df = 6, P=0.0007), industry sector (Deviance = 8.09, Df = 3, p=0.044), job classification (Deviance = 51.81, Df = 11, p=0.0001) and years of experience (Deviance = 51.13, Df = 1, p=0.0001). All other demographic variables showed no evidence of a relationship with the choice of incompetence. The analysis of deviance has shown that while our prediction that years of experience will predict the choice of incompetence is supported, our prediction that qualification will predict the choice of incompetence is not supported. More importantly, the analysis of deviance has revealed other predictors namely age, state, occupation, industry sector, and job classification.

Further, we wanted to know if "how often unethical behaviour occur" predicts the choice of incompetence. Given this predictor is also a categorical variable a GLM was fitted to investigate this relationship. The analysis of deviance has revealed that there is a significant relationship between how often unethical behaviour occur and the choice of incompetence (Deviance = 186.78, Df = 4, p=0.0001) suggesting that this variable is also a predictor for the choice of incompetence.

3.2 The qualitative interviews

3.2.1 Conducting the interviews and analysing the data

The survey was followed by a set of semi-structured in-depth interviews with 43 participants selected from those who responded to the Phase 1 quantitative survey. The interviews were conducted during the month of February 2014 and took place in six Australian state's capital cities. The purpose of these follow-up semi-structured in-depth interviews is to allow for the reporting of participants" perceptions in regards to the nature of the ethical problems experienced in the ICT workplace and how exactly these problems are often solved. All interviews were tape recorded and transcribed verbatim.

Purposive sampling was adopted to select the participants from those who had indicated a willingness to be interviewed. Purposive sampling allowed the researchers to choose cases that were representative of all sub-groups and personal characteristics which might be of interest to the study. The sample drawn included professionals from a range of ICT organisations, both large and small, representing different geographic locations, ages, gender, types of jobs, and employment experience. Table 2 below lists some of the characteristics of the participants whose views have been reported in this study. Participants who are not mentioned in the findings section are not included in Table 2.

Table 2: The characteristics of the participants

Interviewee number	Age	Gender	Years of experience	Occupation	City
2	62	M	43	Project Manager	Perth
3	49	F	32	IT Lecturer	Perth
8	40	M	24	Self-employed	Adelaide
9	40	M	10	Programmer	Adelaide
12	41-45	M	17	Manager	Adelaide
13	67	M	43	Consultant	Brisbane
14	59	M	37	Program Director	Brisbane
15	49	M	25	Business Development Manager	Brisbane
16	43	M	19	IT Manger	Brisbane
17	49	M	16	Business analyst	Brisbane
19	54	M	31	Senior technical specialist	Melbourne
22	55	M	35	Consultant	Melbourne
25	31	M	13	IT Manger	Melbourne

The transcribed interviews were analysed using thematic (qualitative) analyses. Data analysis was completed with the help of QSR NVivo 10, a software package for managing qualitative data. The unit of analysis was each individual interview document. Data analysis proceeded as follows. First, the interview documents were read several times so the researchers could familiarize themselves with the data collected. Next, free nodes (i.e. nodes not organized or grouped) were created based on keywords in the interview documents. Similar text within the interview documents was located and assigned to these nodes. These nodes then acted as "buckets" in the sense that they held all the data related to a specific node. At the end of the creation of the free nodes theses free nodes were further divided into tree nodes. That is, broader categories were developed to group the free nodes. This was to create a hierarchy that made it easy to make sense of the data and facilitate interpretation.

3.2.2 Findings (qualitative interviews)

The quantitative survey highlighted the factors that predicted the choice of incompetence by the survey participants but it did not shed any light on how these participants perceived incompetence. The qualitative in-depth interviews address this issue. Participants framed incompetence in many ways. Table 3. below shows the comments that typify their views. While Table 4 below mentions that 16 interviewees addressed the issue of incompetence during interviews, in Table 3 only the views of 13 participants are included. The reason for this is because the views of the remaining three participants are identical to three views listed in Table 3.

Table 3. Comments that typify participants' views about incompetence

#[2]	Views on incompetence
2	one has nice behaviour and wants to **do the right thing**, the other just perhaps isn't competent so the only way that they can get ahead is by sticking the knives in the back of the people who aren't watching
3	And I have seen some cases of lecturers that clearly were not **up to scratch** eventually being elbowed out.
8	It's like not fixing the problem and just **putting a Band-Aid on it** and then it bleeds and then you've got to put another Band-Aid on it. And then it bleeds and you've got to put another Band-Aid on it.
9	incompetence rises to its highest level - which means like if you're great at your job, you'll get promoted to a new job, and you keep sort of going up the chain until you get to **a job where you're bad at**, and that's where you stay.
12	it starts off from incompetence from **not being qualified** on the system
13	If you know that that's what you're doing then I'd consider that unethical. But **if you don't know**, and I guess it really comes down to how do you validate your own competence
14	I'm talking about are generally highly competent people. But they may **not have the capability** that's needed for that particular task or that particular project.
15	I mean the staff member who's going out might be going out trying to do their best, but they're **technically** incompetent.
16	Personally incompetence when I say to you what is 5 x 4 and someone says **they cannot figure it out** and they can't get hold of a calculator to calculate it then it's incompetence

[2] Interviewee Number

17	So nobody on the project now has done anything wrong but the general aroma of incompetence **has soaked through the whole place** so that things are going to be late and it won't matter much.
19	There is certainly incompetence, but I think it's also **a lack of skills and knowledge**.
22	Incompetence to me is **not knowing how to do the job and not being aware that you don't know**.
25	I don't know that it's necessarily incompetence, but what I think it is that **IT changes an awful lot**. So a technology that, you know five years ago was very mainstream and very popular is becoming less and less popular and new technologies are overtaking it.

The definition that Interviewee 22 gave above, which echoes Interviewee's 13 comment as well, goes beyond the dictionary definition of incompetence i.e. not knowing "how to do" things to highlight the problem of lacking awareness about this "not knowing". As Interviewee 22 puts it: incompetence "is not knowing how to do the job and not being aware that you don't know". As can be seen from Table 3, participants framed incompetence in several different ways but some of the views can be grouped together under fewer themes. The frames "not up to scratch" and "IT changes an awful lot" can be categorised under "not being up-to-date with technology". The frame "lack of skills and knowledge" seems to underpin the frames of "not being qualified", "technically incompetent", "they cannot figure it out", and "not have the capability". Finally, the frames "do the right thing", "putting a Band-Aid on it", "a job where you're bad at" can all arguably come under "Not hiding incompetence".

Table 4 ranks references to incompetence by the occupations of interviewees. It shows that of the 30 occupations represented in the 43 interviews, there were a total of 48 references to issues of incompetence. As a single group, consultants had more to say on the topic than any other occupation (20.8% or 10 of the 48 references). However, self-descriptions of occupation reveal that it is possible to generalise roles. For instance, various people interviewed had a management role, although their actual occupation titles differed. These included project manager, IT manager, business development manager, senior project manager, manager, program director, national instructor manager, managing director, chief information officer, program manager and operational manager. When ones abstracts occupation to role, managers accounted for 47.9% or 23 of the 48 references. In terms of the number of interviewees, Table 4 shows that managers accounted for 43.8% or 7 out of 16 interviewees who addressed the issue of incompetence in the ICT workplace.

Table 4. Incompetence by interviewee occupation

Occupation	References	Interviewees
Consultant	10	2
Project Manager	7	1
IT Manger	5	2
Business Development Manager	5	1
IT Lecturer	3	1
Programmer	3	1
Senior Software Engineer	3	1
Senior project manager	3	1
Self-employed	2	1
Manager	2	1
Business analyst	2	1
Phone Systems Installer	1	1
Program Director	1	1
Senior technical specialist	1	1
Technical Writer	0	0
Data Analyst	0	0
Chief Operating Officer	0	0
National instructor manager	0	0
Business owner	0	0
Database/IT coordinator	0	0
Managing Director	0	0
Graduate Business Analyst	0	0
CIO	0	0
Program Manger	0	0
ICT policy and reporting	0	0
Public servant	0	0
Tester	0	0
Accreditor	0	0
Senior Enterprise Architect	0	0
Operational Manager	0	0

Although incompetence in the literature was strongly linked to dependability and trust, dependability was not something that arose in the interview data. 24 Interviewees did refer to issues of trust, however a NVIVO matrix query revealed that none of the references to trust corresponded to references of incompetence.

The survey results also revealed that with increasing years of service, issues of incompetence were recognised. Although actual age, rather than an age range, was used in the interviewees, this was also borne out in the interview data. No one with less than 10 years experience mentioned issues of incompetence, only one person with 10 years and one with 13 years, then from 16 years of experience onwards, incompetence became more of an issue for interviewees.

3.3 The focus group interviews

3.3.1 Conducting the focus group interviews and analysing the data

In this phase the findings from the second phase were used to provide the input for focused discussions with key participants regarding effective strategies to help ICT professionals solve problems such as incompetence in the workspace ethically. The purpose of these focus group interviews was to determine which were the effective strategies for overcoming ethical problems that could be developed into a web resource available to all ICT professionals. A set of five focus group interviews were conducted with key informants in major state's capital cities in Australia. Each focus group discussion comprised participants who were senior professionals, with at least two ACS Fellows in each group (distinguished, senior members of the ACS). Focus group interviews are very appropriate when the interest is in understanding an idea, opinion or an experience, when the topic is impersonal enough to be posed to a group, and when it is important that the participants integrate other participants' views into their responses or build upon them. While purposive sampling was again adopted here, the commonality of the

participants background (certainly not their opinions) and their geographic locations were the most important factors in the selection of participants. All the focus groups were tape recorded and transcribed verbatim. The transcribed focus groups were analysed using thematic (qualitative) analysis in the same way the individual interviews with the ICT professionals were analysed (see data analysis section above), but in this case the unit of analysis was not each individual who participated in the focus groups but rather the unit of analysis was the whole conversation. Data analysis for this phase was also completed with the help of QSR NVivo 10.

3.3.2 Findings (focus group interviews)

The views about incompetence varied between focus group participants, as it had for interviewees. For instance, one participant from the Canberra focus group distinguished technical and social competence, giving examples of people who were technically capable, yet socially inapt, and that such social incompetence affected their workplace performance, even though they were technically competent.

Incompetence was related to management. A Canberra participant stated that "*the competence levels of project managers is just woefully low*". Similarly a Melbourne participant reflected that "*I believe that is a generic IT issue now. I mean, I've been in IT from 1990's onwards, and the competence has been dropping, particularly of the senior managers, they've dropped shockingly."*

Solutions to incompetence were seen in education. A Canberra participant said "*My answer to the incompetence story is mostly education*". A male Melbourne participant claimed: "*When you see the word 'incompetence' as an ethical problem, this tells you about the company culture, it's not one that supported of continuous learning*". A female Melbourne participant countered with "*I was going to agree with the point that <name> made earlier on about if you're dealing with incompetence in another person, or a person that you're working with and that goes back to the organisational culture ... this industry changes at an incredible rate, you have to actively learn and develop your skills continuously and 6 months of not paying attention to whatever is on the leading edge and you're in trouble, basically*". A Sydney participant had a similar view, but also linked this discussion to trust: "*I think our certification could be part of this solution, get, looking down there the incompetence, lack of knowledge, trust, lack of skills, lack of qualifications, I see things grouping there*". Another Sydney participant claimed that an effective strategy for addressing issues of incompetence might be through industry standards, such as SIFIA; that was from a Business Development Manager with 25 years ICT experience, and makes sense, in that a person recognised to operate at a given SIFIA level has competence to work at that level, but not at a higher level.

Another male Melbourne participant raised the point that "*one thing which we are ignoring is a work pressure here. Okay, if a senior manager might be a senior manager who is very much competent, but the way it's been structured, and he's having too many eggs, you know, in one basket. So that's making him incompetent, though he can perform very well, but just because of the cost-cutting or just because you know, to achieve some goal, he's been given too many* tasks which is making him incompetent to perform or give his best". It appears the findings from the focus group interviews are consistent with the findings from the survey and the qualitative interviews in that the experienced professionals have a greater concern about incompetence than others.

4. DISCUSSION AND CONCLUSION

This project collected data not only on the most common ethical problems experienced by ICT professionals in Australia but also on the approaches that should be employed in solving these ethical problems. The data collection for this project included conducting a quantitative survey, conducting qualitative interviews and conducting focus group discussions. This complex data gathering plan using the methods outlined above yielded good data on all the information being sought during all the three phases of the project. The methods related to each other in this way: the quantitative survey identified incompetence as the fifth from the top in the list of the most common ethical problems experienced by ICT professionals. The findings from the survey helped shape both the content of the follow-up interviews and the selection of those interviewed. The findings from the qualitative interviews provide detailed and rich accounts about how ICT professionals view incompetence in the Australian ICT workplace. The findings from the qualitative interviews were the input for the focus groups discussions.

Of the 2,315 respondents who participated in the survey, the MRF analysis revealed that incompetence was ranked fifth from the top of a list of the 57 most common ethical problems experienced by ICT professionals (N=750, 35.9%). An inspection of the results of the cross tabulations revealed that 34.8% described their occupational category as manager and 29.1% indicated they were consultants. The GLM has found a significant relationship between the choice of incompetence and occupation (Deviance = 23.15, Df = 6, P=0.0007) suggesting occupation, among other things, does predict the choice of incompetence. The findings from the qualitative interviews are consistent with the above findings. A cross referencing of the interviewees responses that addressed the issue of incompetence during the interviews against their occupation has revealed that consultants had more to say on the topic than any other occupation (20.8% or 10 of the 48 references). This is followed by managers who accounted for 14.6% (7 of the 48 references). These findings indicate that the experienced professionals have a greater concern about incompetence than others; an observation that the findings from the focus group interviews have also confirmed. Obtaining such findings would not have been possible had only one method been used.

From an ethics perspective the most interesting finding comes from the qualitative research. While it is significant that more experienced professionals have a greater concern about incompetence, as found in the quantitative research, this is perhaps understandable. More experienced and older workers often believe that standards are slipping and that things "were better in the old days". What is of more significance is what is said about incompetence in the interviews and focus groups. Incompetence is commonly considered a moral failing of an individual. It certainly is this but the individuals may not be totally to blame if incompetence is at least partly a function of the positions into which they are required to operate. The moral failing extends to the management that expects people to work competently in situations where this is not possible or extremely difficult, either because they are not adequately trained in that area or they are expected to do more than anyone can competently do. The ethical concerns raised by the research point not just to the professionals doing the actual work but to the management and perhaps to the organisation as a whole.

As ICT becomes ever more pervasive in society, through mobile, virtual, cloud and other technologies, it becomes increasingly

more difficult for ordinary citizens to be able to judge the quality of ICT products. The public has to be able to trust in the competence of the people developing those products and that in turn calls for greater degrees of professionalism in the ICT workplace, not only of the professionals themselves but also of the organisations. Later work will seek to inform the tertiary community about the real challenges and strategies for solving ethical problems that are experienced in the ICT workplace, so that in future their graduates can be better prepared for the types of challenges they will face.

5. ACKNOWLEDGMENTS

The research reported here is supported by an Australian Research Council Linkage grant (LP130100808), for which the industry partner is the ACS. Another team investigator was Mr John Ridge. The authors also wish to thank Sharon Neilsen, Director of Quantitative Consulting Unit at Charles Sturt University, for her help with the statistical analyses used in this study and Professor Ken Russell, also from Charles Sturt University, for his valuable comments on the results from these statistical analyses.

6. REFERENCES

[1] Genus, A. 1997. Unstructuring incompetence: Problems of contracting, trust and the development of the channel tunnel. *Technology Analysis & Strategic Management.* 9, 4, 419-436. Retrieved from http://search.proquest.com/docview/226899027?accountid=10344

[2] Beckstead, J.W. 2005. Reporting peer wrongdoing in the healthcare profession: the role of incompetence and substance abuse information. *International Journal of Nursing Studies.* 42, 3 (March 2005), 325-331. DOI= http://dx.doi.org/10.1016/j.ijnurstu.2004.07.003

[3] Avizienis, A.; Laprie, J.-C.; Randell, B.; Landwehr, C. 2004. Basic concepts and taxonomy of dependable and secure computing. *Dependable and Secure Computing, IEEE Transactions.* 1, 1 (Jan.-March 2004), 11-33. DOI= 10.1109/TDSC.2004.2. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1335465&isnumber=29463

[4] Glass, R.L. 2013. The Queensland Health Payroll Debacle. *Information Systems Management.* 30, 1, 89-90, DOI= 10.1080/10580530.2013.739899

[5] Baba, M. 1999 Dangerous Liaisons: Trust, Distrust, and Information Technology in American Work Organizations. *Human Organization.* 58, 3 (September 1999), 331-346. DOI= http://dx.doi.org/10.17730/humo.58.3.ht622pk6l41l35m1

[6] Euchner, J. A. 2011. Innovation's "Skilled Incompetence". *Research Technology Management*, 54, 5, 10-11. Retrieved from http://search.proquest.com/docview/894748350?accountid=10344

[7] Merriam Webster Dictionary. 2013, Trust (verb). *Merriam-Webster Dictionary*. Retrieved on June 1, 2013 from http://www.merriam-webster.com/dictionary/trust.

[8] Good, D. 2000. Individuals, Interpersonal Relations, and Trust, in Gambetta, Diego (Ed.) *Trust: Making and Breaking Cooperative Relations*, electronic edition, Department of Sociology, University of Oxford, chapter 3, pp. 31-48.

[9] Rempel, J.K., Holmes, J.G., and Zanna, M.P. 1985. Trust in Close Relationships. *Journal of Personality and Social Psychology.* 49, 95-112.

[10] Weckert, J. 2005. Trust in Cyberspace, in R. Cavalier, (Ed), *The Impact of the Internet on Our Moral Lives*, State University of New York Press, New York, pp. 95-117.

[11] Al-Saggaf, Y. & Burmeister, O.K. 2013. A survey of Australian ICT professionals' perceptions regarding the most common ethical problems they face in the workplace. In M. Warren (Eds.). *Proceedings of the Seventh AICE Conference* (pp. 43-48). RMIT, Melbourne, Australia. December 3, 2013. Retrieved from http://auscomputerethics.files.wordpress.com/2014/02/aice-proceedings-2013.pdf.

Twenty-five Years of ICT and Society: Codes of Ethics and Cloud Computing

Diane Whitehouse
Business Partner
The Castlegate Consultancy
diane.whitehouse@thecastlegateconsultancy.com

Penny Duquenoy
Principal Lecturer and Researcher
School of Science and Technology
Middlesex University
p.duquenoy@mdx.ac.uk

Kai K. Kimppa
Postdoctoral Researcher
Information System Science
University of Turku
kai.kimppa@utu.fi

Oliver K. Burmeister
Associate Professor
School of Computing and Mathematics
Charles Sturt University
oburmeister@csu.edu.au

Don Gotterbarn
ACM Committee on Professional Ethics
gotterbarn@ACM.org

David Kreps
Senior Lecturer
Salford Business School
University of Salford
D.G.Kreps@salford.ac.uk

Norberto Patrignani
Associate Professor
Doctoral School
Politecnico di Torino
norberto.patrignani@polito.it

ABSTRACT
Celebrating achievements is an important social ritual. Tracks and themes at conferences such as ETHICOMP 2015 provide opportunities for the careful discussion of challenges facing society in terms of information and communication technology (ICT). This topic provides the underpinning rationale to the body of papers presented throughout the entire ICT and Society track at this ETHICOMP conference. The conference orientation explains this paper's focus on codes of ethics, professional ethics, organisations and the particular challenges of the cloud and virtuality over a 25-year time-period.

Categories and Subject Descriptors
K4 [**Computers and Society**]

K7 [**The Computing Profession**]

K4.1 [**Computers and Society**]: Public Policy Issues – ethics, regulation

K4.2 [**Computers and Society**]: Social issues

K4.3 [**Organizational impacts**]

K7.4 [**Professional ethics**]: Codes of ethics; codes of good practice; ethical dilemmas.

Keywords
Codes, cloud computing, ethics, society, virtuality.

1. INTRODUCTION
The International Federation for Information Processing (IFIP) is an umbrella society that encompasses more than 50 information and communication technology (ICT) societies globally

IFIP has a long history of focus on work on ICT and Society. The federation has conducted a more than 25-year international exploration of how ethics, and especially professional ethics, can operate in national computing associations. In this regard, it has preoccupations that are similar to several other international professional computing organisations. Operationalising such preoccupations can at times be quite challenging, as this paper indicates. Yet, progress and important discoveries can still be made, and lessons learned can certainly lead to an improvement in focus.

This paper first examines the general development of codes in three associations, with a specific concentration on the efforts made to develop a code of ethics in IFIP (this has resulted in the creation of a discussion space called Special Interest Group 9.2.2 on the Framework of Ethics of Computing and a parallel focus on professional ethics). It shows that on-going work which manifests a concern for the ethical and societal implications of computing use in contemporary society is important (as can be seen through IFIP's 2014-created domain committee on cloud computing and its more established work on virtuality). Last but not least, the paper examines possible ways forward with regard to an IFIP code and explores briefly arguments in favour of a business case for an ethics of computing.

2. CODES IN ICT SOCIETIES AND ASSOCIATIONS

Professionals often make public statements on how they understand their ethical responsibilities and pledge to meet them: they call these statements codes. The different function of codes affects their content. A first type is called a code of practice (or occasionally conduct), and a second is called a code of ethics [3]. A code of conduct is about the personal conduct of an individual professional. A code of practice is thus about practices manifest in the profession, that go beyond the behaviour of the individual. The practices include how an organisation deals with its vendors or the types of methodologies in which it engages. Codes of practice contain significant detail so as to facilitate easy detection of code violations; codes of ethics include more general statements of the profession's moral values used in practice to guide their practical decisions. IFIP has adopted softly this distinction [14], but has occasionally merged these two code functions, as it did in its 2011 code of ethics checklist which it called a "disciplinary code" [13].

2.1 Association for Computing Machinery

The early examples of computing codes were codes of practice. The Association for Computing Machinery (ACM)'s 1978 Code [2] was a list of specific imperatives for a slow-changing profession, but it then fell out of step when newer technologies came on board. Since computer professionals began to need help in identifying and addressing ethical issues, in response, societies shifted their codes towards an emphasis about the moral guiding aspirations of the computer profession, and included statements on these. In the 1990s, they moved their codes from being regulatory codes to more normative codes [11].

In 1992, ACM's representative to IFIP's Technical Committee 9 on ICT and Society, led the revision of the 1978 ACM Code of Ethics. The code was to be based on the following eight unchanging moral imperatives: Contribute to society and human well-being; Avoid harm to others; Be honest and trustworthy; Be fair and take action not to discriminate; Honour property rights including copyrights and patents; Give proper credit for intellectual property; Respect the privacy of others, and Honour confidentiality. Each imperative was illustrated by a set of clauses that showed how the particular constant imperative applied to the changing technology. This code used extensively "the draft IFIP Code of Ethics, especially its sections on organizational ethics and international issues."[2].

Early IFIP discussions about the possibility for a universal computing code emphasised the difference among cultures (i.e., their multiculturalism) rather than the common elements between professionals in different cultures (inter-culturalism) [9], and their agreements about acceptable professional behaviour. This multi-cultural focus contributed to IFIP's reticence about the possibility that existed to establish a single code of ethics. This cautiousness, however, had the positive effect of contributing to the development of a support group (hereafter called SIG 9.2.2, a Special Interest Group on a Framework on Ethics of Computing). Its mandate was to provide "spaces for discussion" [7], places where members of the computer community could discuss common issues of ethical concern. This development and the role of the special interest group is described in more detail in section 4 of this paper.

One element said to distinguish professionals from other practitioners is that the former pledge themselves to certain moral responsibilities and a higher order of care. IFIP embraced this distinction in its standard for certifying an organisation's professional code of conduct [13]. In earlier codes of practice, a deontological approach to ethics was pursued: a person's competent completion of a particular specification was considered sufficient for him or her to be named as an ethical professional. Such an approach is no longer held to be adequate for a society's code of ethics to be certified by IFIP. In its code of ethics assessment standards, IFIP now requires consideration to be taken of the impacts that arise from any act or action on a broad range of stakeholders, not just those who have a financial interest in the computing product delivered [13].

2.2 ACM/Institute of Electrical and Electronics Engineers Computer Society

In the late 1990s, the ACM/Institute of Electrical and Electronics Engineers Computer Society (IEEE-CS) Software Engineering Code of Ethics and Professional Practice version 5.2 used the concept of discussion spaces when developing its code that was designed for a profession rather than for a particular national professional society. The code was put together by a multinational taskforce with inputs from other practicing professionals, and representatives from industry, governments, the military, and education.

An electronic discussion space was used to identify the ethical imperatives to be included in the code. The ACM and the IEEE-CS appointed joint chairs to manage this effort. The remainder of the task force was composed of volunteers who responded to an international call for participation sent out to numerous professional computing societies and posted on bulletin boards and in academic and practitioner computing publications. Suggestions and comments from practitioners, members of professional societies and academics, who were all related to software engineering, were solicited. This multi-national group developed the imperatives which were refined during the code's development. The code went through several reviews as it moved toward consensus-building in the electronic discussion space. Most clauses in the proposed code received a better than 90% approval rating in open voting.

Since its initial approval in 1998, the code has been translated into multiple languages. Its content has been adopted by professional societies across the globe. The need for such localisation was a challenge that IFIP's individual members were already trying to make clear in their own initial position-taking in the mid-1990s.

This adoption indicates both that the code itself captured the conscience of a profession and that the development of a universal code of ethics is possible. First, the Software Engineering Code asserts that, when there is a difficult ethical decision to be made, it is the welfare of the public that is the overriding consideration – this assertion supports the 2011 IFIP position on the consideration of public impacts [13]. Second, the Software Engineering Code received a consistently high level of agreement about the behaviour expected of professional software engineers. Unfortunately, a clear understanding of these ethical obligations does not guarantee that they will be observed by every software practitioner. In addition, one must remember that the original clauses in the code were not accepted in consensus, and thus differing views on these matters existed originally and still do exist.

Perhaps the question of enforcement of these standards can therefore be considered as a local issue.

3. AN INTERNATIONAL-LOCALISATION HYBRID?

For IFIP, over a 25-year period, attempts to achieve a global code have faced several difficulties: the rapidity of technological change; cultural diversity; and the tension that exists between the fact that codes require both specificity, so as to hold members accountable to their details, and generalisation, without being too overly general that they are not of concrete use. A code needs to be general enough so that it does not require continual updating as technologies change, yet it must be specific enough that it is possible to hold its members accountable in terms of their behaviours. Typically changes in technology at some point require modifications in professional behaviour.

3.1 IFIP History

Twenty-five years ago, in IFIP, a newly-formed task group raised these several challenges with the federation's member societies. The task group sought feedback from the federation's member ICT societies about how rapidly a computing society code of ethics should be updated, as new technologies emerge [6]. More details on this approach are tackled in section 4 of this paper.

In IFIP, a values approach for a global code of ethics was first discussed in terms of commonly agreed principles. The seed for that approach can be seen as early as 1996, as Jan Holvast proposed [12]. Holvast suggested that – if IFIP had general principles that were acceptable to all its member national societies – this generalisation might provide a means to making progress towards a global discussion. Originally through its Taskforce on Ethics, which then became Special Interest Group 9.2.2 on the Framework of Ethics of Computing, IFIP played an influential role in exploring what a global code of ethics should be.

IFIP began work on a global ICT code of ethics in 1988, under the leadership of Harold Sackman. At times, commentators have taken the view that such a code would be impossible to establish, as did IFIP for a time. It announced: "For an international organization like IFIP, formulating a code acceptable for all Members will be an impossible task" [12]. And, "To conclude, SIG9.2.2 does not believe that a universal or international code of ethics can be mandated for persons working with information and communication technologies" [5].

IFIP SIG9.2.2 – with its focus on a Framework of Ethics of Computing – is the successor to that earlier task force: "The mandate of the IFIP General Assembly that created that Task Group in 1992 was to explore the feasibility of an IFIP worldwide code" [4]. Over the years, SIG9.2.2, and its predecessor task force, have held various events and meetings, and produced a series of publications to which representatives many national ICT societies have contributed.

3.2 Codes and Their Complexity

The independent development of codes of ethics in ICT societies has led to codes being referred to by many names. While there are codes of conduct, codes of ethics, codes of practice, and codes of professionalism, there are also many more titles.

A common thread identified among codes, however, is their two-part structure. Codes tend to consist of first, a set of values followed, second, by implementations provided by a normative professional practice. Some computer societies have used a model for a code which consists of a high-level set of aspirational imperatives: in text under that imperative, they offer more specific illustrative examples of how those imperatives apply. The idea is that the ethical imperatives are constant. Therefore, the only modifications needed are in the example clauses. (There is, nevertheless, a practical problem to the modification of codes if the codes are translated into many different languages. Having one version of a code, and differing translations, can cause confusion.)

Which part of the code is most often subject to change? It is the second, professional practice component. The first parts, i.e., the underlying values that govern practice, only rarely require changes in response to technological changes. It is the way in which those values are to be applied that can alter more often. Achieving global agreement on a code of ethics entails, on the one hand, many logistics as well as impracticalities and, on the other hand, the bringing together of many stakeholders.

It is therefore best to make changes to such texts as infrequently as possible. Thus, it would appear that – in a two-part code of ethics structure – the first part lends itself more effectively to stakeholders being able to reach (international) agreement about it than the second. A practical way forward in making progress may be located in this two-fold description of codes of ethics.

3.3 A Way Forward?

The lesson emerging from 25 years of IFIP activity, of partial successes to derive a globally agreed set of professional practices, suggests that professional practice needs to have localised interpretations.

The set of professional responsibilities that arise from values should be subject to culturally diverse interpretations and, in turn, agreement on what general principles – or values – can be turned into concrete statements, against which members can be held accountable. Separating practice from values enables the second, localised, practice part of a code of ethics to be modified in ways that are culturally appropriate to technological changes, while leaving the internationally agreed values unchanged.

A hybrid solution to these challenges was proposed by IFIP in a 2012 Malaysian conference [8], to which the IFIP President, Leon Strous, sent his representative, Oliver K. Burmeister. The paper [8] presented four examples of the value, honesty. It showed that honesty is a value that is common to four ICT societies, on four separate continents. However, the professional practice related to the value of honesty is interpreted differently in those four countries. Thus, while it may be possible for all ICT professional societies to agree on such a value, the way in which a particular value is practiced can differ inter-culturally and intra-culturally over time.

An international-localisation hybrid code of ethics can value both harmony and diversity, and can permit the tension between these two values to exist in a less problematic fashion. There may still, however, be debate and dialogue about the subtle ways in which the application of the values reflect more precisely the higher-level values.

Therefore, one possible way forward would be to conclude a limited global agreement while, at the same time, appreciating the importance of diversity in professional practice. This international-localisation hybrid would consist of yet-to-be-determined common or universal values, which are then implemented locally in the practice of the profession. Unlike prior attempts at complete harmony or, conversely, attempts to criticise the creation of a global code of ethics because of the many differences that exist between ICT societies, this international-

localisation hybrid approach could provide a means of making progress on discussions about a global code of ethics – an approach that warrants further investigation.

4. SOME EXAMPLE ACTIVITIES

Here, three ways which IFIP has developed to tackle ethical debates are described. The first is IFIP's Special Interest Group on a Framework on Ethics of Computing, the second is IFIP's newly established Domain Committee on Cloud Computing; and the third is one of the federation's working groups on virtuality. Each is explored at different levels of detail.

IFIP handles such issues by using several approaches: increasingly, it is developing cross-domain activities that bring together diversities of stakeholders (the Domain Committee on Cloud Computing is but one example); it also organises dialogue on a variety of societal issues (examples include digital equity and the relationship between ICT, society and education/training), particularly through the comprehensive work of its Technical Committee 9 on ICT and Society.

Following the descriptions offered in these three example areas, the authors of this paper return to the on-going challenges that face the actual formulation and implementation of professional computing codes of ethics.

4.1 A Framework on Ethics of Computing

In the early 1990s, an IFIP task group was set up with the goal of clarifying "how to handle the question of Ethics within IFIP". It took into account "the way the national Societies, Technical Committees and Working Groups viewed the question" [7].

During this initiative, an analysis of codes from member societies and other relevant professional bodies was undertaken to assess the codes' themes and their approaches to ethics with the view of drawing together a set of commonalities from which a set of IFIP guidelines could be derived.

Different traditions among the IFIP member organisations from a wide range of countries meant that there were quite different approaches to codes of ethics. A typical example is the different way of looking at and applying codes of ethics. On the one hand, European "continental" computing societies, such as the German Gesellschaft für Informatik e.V (GI) or the Finnish Information Processing Association (FIPA/TIVIA) tend to view codes of ethics rather as suggestions or reminders to their membership about what is ethical and what is not. When there is a need for more regulations, the continental approach relies on laws that are passed rather than self-regulation. Codes in the United States of America and the United Kingdom (UK), on the other hand, provide examples of a more strict adherence to the code of ethics; even the format of the code emphasises its binding nature on the organisation's members. For example, in the UK code the members "shall" or "shall not" do certain things.

The culmination of the original work of the task group was published in 1996 under the title *Ethics of Computing: Codes, spaces for discussion and law* [7]. Its first section comprised five chapters on themes around codes of ethics and professionalism relevant to IFIP and its member societies. The conclusion from the work undertaken was that the "main task inside IFIP is to create "spaces for discussion" on ethical questions. The previously mentioned Special Interest Group on a Framework on Ethics of Computing was created to further this remit. The group set up round-table sessions in 1998 at the IFIP Human Choice and Computers conference to discuss the ethical issues that were of concern to the professional community at that time. The results of the workshop were published in the monograph *Ethics and the Governance of the Internet* [6].

The special interest group's next task was to take into account the different approaches to the expression of codes of ethics, and to develop a booklet citing the *Criteria and Procedures for Developing Codes of Ethics or of Conduct* [5] so as to aid national societies to create their own codes. The document does not mandate any topics to be part of a code, but offers suggestions on topics that should be considered in a code of ethics.

IFIP's international perspective is important when exploring and discussing ethical issues, as is the federation's link with professional computing societies. IFIP exists to provide its members – which are national computing societies – a common international forum. It is therefore important to provide input on ethics to the people training future ICT professionals, to support current ICT professionals, and contribute to policymaking. Many of the individual members of the various IFIP working groups and technical committees are academics. A considerable number of them are in practice engaged in research projects working with industry and ICT practitioners (to provide ethics input on advisory boards) or are working on projects to embed ethics and ethical reflection into technology development. Others work in industry or commerce or in standardisation bodies. Many of the federation's individual members are nominated representatives of their national computer societies or have strong links with their national society.

Ethics is at the heart of professionalism, in whatever field, in terms of maintaining the standards of the profession and providing support and guidance for members working in it. In line with the earlier approach of helping national societies with their own ethics governance, the group therefore produced a second document to aid national computing societies in creating ethics groups (or committees), which could then tackle any ethics questions raised by the national societies – one of those being whether a code of ethics was needed, and if so, of what kind [4].

4.1.1 Approach and Activities

Creating spaces for discussion on ethical questions relating to ICT is the foundation of the work programme of this special interest group. As has been seen, its work began with a series of workshops held at the IFIP Technical Committee 9 Human Choice and Computers Conference in 1998 on the ethical issues of the day. It has been maintained over the years through participation in IFIP TC9 conferences, joint initiatives with IFIP's Working Group 9.2 on social accountability and ICT including the regular, annual IFIP summer school series on privacy and identity http://www.ifip-summerschool.org, and meetings of its own.

Members of the special interest group bring together a variety of perspectives to the group's work. The group is informed by the members' very different areas of expertise, such as education, health, human-computer interaction, security and software engineering, applied ethics, and ethics in governance of ICT including professionalism and codes of conduct. With this broad mix of skills, the group has been well placed to provide IFIP with intellectual input on contemporary ethical challenges.

Most recently, in February 2015, a London-based workshop was organised on "The challenges of virtuality and the cloud" http://ifipwg92.org. This workshop came about as a result of a joint initiative between the special interest group, Working Group 9.2 on social accountability and computing, and the British

Computing Society (now known just by its acronym, BCS) ICT Ethics Specialist Group.

While the activity of this special interest group emerged directly as a result of a computing association's concern for ethics and the extent to which a code encompassing it should be created and designed, many of the federation's other activities are derived from a concern with contemporary ICT challenges.

4.2 A Domain Committee on Cloud Computing

A core subject for contemporary exploration in these terms is cloud computing. It is potentially one of the main paradigm shifts in the recent history of computing and information processing, and poses considerable societal and ethical challenges.

The IFIP domain committee on cloud computing was established in September 2014 by IFIP's President, Leon Strous. Its aim is to join together efforts from all of IFIP's technical committees and working groups to explore the domain of cloud computing. Its first meeting took place in the Austrian capital of Vienna, in September 2014; a second meeting was held in Beijing, China in June 2015. The committee's first deliverable is to be an IFIP position paper on cloud computing, and a policy statement that will be presented at the next IFIP General Assembly in October 2015 in Daejon, Korea [1].

The cross-cutting character of the committee underlines the importance of taking a transversal view to cloud computing through the creation of a horizontal approach that spans the whole of IFIP. It permits the development of a human-centred perspective on cloud computing – computing for the public interest that includes the cloud's potential impact on public authorities, small- and medium-sized enterprises and society at large, including the realm of science. This enables IFIP to speak with a single voice on this important topic and to establish collaboration with other international organisations such as the ACM Special Interest Group in Computing and Society (ACM-SIGCAS) and the IEEE-CS.

IFIP's domain committee on cloud computing aims at addressing the societal and ethical challenges of cloud computing. It does so by providing analysis, recommendations and suggestions of possible implications to different forms of stakeholder. The stakeholders are composed of companies (including small- and medium-sized enterprises as well as large corporations), society (including public authorities, users, scientific community), and policymakers. The committee also addresses the cloud's implications from the sustainability and the environmental points of view, its possible connections at a global scale, and the need for a human-centred view of the cloud.

4.2.1 Cloud Computing and Its Challenges

There are many different views of cloud computing, some advantageous, some less beneficial – all are being explored by the domain committee. There are at least six different perspectives that can be outlined.

From a particular perspective, the promises of cloud computing – as a global infrastructure that is network-based, multi-tenant, scalable and self-service – is very attractive for many organisations, in particular for small- and medium-sized enterprises and for those in the public sector. Organisationally, the cloud represents a big shift in reverse, back towards centralised architectures: end-users and organisations will consume on-demand resources provided by very large data centres. As a consequence, many chief information officers will have to redefine their ICT governance role inside their organisations. From the societal point of view, the independent status of digital citizens that was provided by the distributed and personal computing of the 1980s, through to the early part of this decade, risks being lost as people simply become digital consumers. If the domain committee were to take a Science, Technology and Society view that technology and society co-shape each other, and were to start by looking at cloud computing as a socio-technical system, what kind of society might be shaped by this new ICT direction? From a human-centred view of cloud computing: an evolution in this direction identifies considerable opportunities: they include the cloud's always-on capability, and the access it provides to storage and processing power anywhere, anytime. Yet, on another front, the accumulation of personal data at global scale in the cloud's big data repositories introduces a dimension of risks for privacy on a scale that will require international policies and norms to be established. Other important issues include the facts that: traditional borders of organisations will disappear (this is called organisational de-perimeterisation); there is a need for precise definitions of the role and contractual obligations of cloud providers and ICT-as-a-service is now available. Other challenges remain around: risk management; legal status; compliance; data availability and ownership; the risks of lock-in; and privacy and security.

4.2.2 Challenges to Stakeholders

Important implications of the cloud arise for many stakeholders. The opportunities that arise for two stakeholders, the scientific community and policymakers, are explored here alongside one particular challenge, that of the environment which is also in its own right a stakeholder.

4.2.2.1 The Scientific Community

For scientific applications, IFIP could support, for example, at a Global Computing Scientific Cloud through an effort such as the European Organization for Nuclear Research (CERN) in Physics, based in Geneva, Switzerland. It could focus on a Cloud Computing for Science initiative, open to researchers from all over the world, and so would avoid the commercial take-over of computational science.

4.2.2.2 Policymakers

Many national authorities are encouraging their agencies to adopt cloud computing for their ICT services. While this provides them with interesting opportunities for saving costs, coordination is required at the policy and standardisation levels.

4.2.2.3 The Environment

ICT use can, of course, optimise the processes of de-materialisation, and reduce pollution. However, in the case of cloud, the power consumption of gigantic data centres owned by cloud providers should be carefully taken into account. On the one hand, emissions due to ICT are estimated to reach 1.25 gigatonnes of CO_2 ($GtCO_2$) by 2030 (28.8% due to data centres, 47.2% to end-user devices, and 24.0% to networks). On the other hand, also by 2030, the CO_2 reduction caused by a wise use of ICT (including functional optimisation and dematerialisation whether in mobility, manufacturing, agriculture, buildings, or the energy, sector) could reach 12.08 $GtCO_2$ [10].

While the balance appears to be positive, it would be more appropriate to take into account the entire life-cycle of ICT in the calculation of the CO_2 emissions that will be due to the increased

manufacturing and development of ICT, and the related growing problem of e-waste management.

As always at global scale, it is also important to analyse the consequences of the concentration of storage and computing power in certain countries and continents, while other countries and continents become progressively more and more dependent on the first for their ICT services. Could this growing concentration of ICT power create a drive towards a form of cultural imperialism, and difficulties in dealing with diversity?

These are just three specific examples among all the societal and ethical questions that face cloud computing, as an example of the questions surrounding particular domains of contemporary ICT development. They raise a very practical set of challenges that are best handled by an amalgamation of researchers and scholars, scientists, policymakers, standardisation bodies, and industrialists working together. Last but not least, national and international professional computing bodies have clearly a considerable role to play in such an initiative.

4.3 A Working Group on Issues of Virtuality, Society and Ethics

IFIP's Working Group 9.5 on Virtuality and Society, which was established in 1989, has focused its interests on the increasingly integral nature of ICT not just to people's daily lives but also to their understanding of themselves.

The notion of virtuality can be usefully looked at through the lens of the evolution of the Web in its different stages of 1.0, 2.0 and now 3.0. In today's Web 3.0, the Internet of Things, even the notion of users and developers is problematised: the user becomes or contributes data; users are no longer confined to a Web 1.0 passivity or are merely the labourers and tools for the generation of content within Web 2.0 social networking. The celebrated phenomena of reblogging and retweeting, and of being part of a crowd from which data is sourced, turns users into channels – the cogs of a machine – so that they become part of the network and elements in a wider application.

In short, the virtuality of people's engagement with ICT has increased, or become more intense: this, as the web has moved through different stages in its development, and that engagement has become ever more integral to its working.

Both people's theoretical understandings, and an ethics that is able to withstand the intensity of such engagement, must not only be robust but must be capable of keeping up with developments, as the inevitable next shifts in direction begin to loom.

5. TOWARDS A BUSINESS CASE FOR ETHICS

Governance – i.e., codes, policy development, corporate or organisational behaviour, individual behaviour, and standards – are therefore of increasing importance.

Historically, various international and national societies have been involved in the development of many of the documents referred to in this paper, and ICT societies' members were consulted on the development of many of the procedures outlined. Stories of the creation of national codes of ethics or conduct were collected so as to learn from actual ethics committees' examples of creating codes.

The next 20 years are bound to change the whole computing and organisational landscape considerably In Europe, purely as an example, the Council of European Professional Informatics Societies (CEPIS) is now aiming to create a Europe-wide code; however, it is running into the same challenges in cultural differences that IFIP faced two decades ago. This is clearly not surprising, since although the European Union is a union, it consists of 28 different member states, with often very different approaches to computing and organisations: while some key issues have been legislated on, many others are subject to the principle of subsidiarity and lie within the purview of decision-making by individual nations. In particular, since 2004 and 2007, there are the 12 new member states, many from eastern and southern Europe, that seem to express a rather pragmatic view at the ethics of ICT. This could be described as being along the lines of, "We'll deal with that when we have our businesses up and running".

ICT is becoming more and more international. It would seem that the work of national computing societies cannot be the only solution; rather, that an international code is needed. Hopefully, initiatives such as Value Sensitive Design and Responsible Research and Innovation – currently promoted by the European Union's Horizon 2020 calls for research proposals http://ec.europa.eu/programmes/horizon2020/ can enhance the situation not only in Europe but also internationally, since ensuing projects can include international partners.

Policies too can be influenced. They can be developed in all sorts of areas – from company policy in general to more specific fields, such as cloud computing. Part of any policy is to do with governing behaviour: this is an area of ethics, namely professional conduct.

A business case can therefore be made that, through ICT ethics, policy, governance and workplace behaviour generally are all being influenced. This, in turn, ultimately leads to the conclusion that ethics is about helping people.

Ethics related to ICT is well worth pursuing, since it can:

- Produce better services for clients.
- Produce better products.
- Make the goods and services produced competitive vis-à-vis other products or services, since they include some in-built element of guarantee about the standards on which they are based.
- Raise the standard of professionalism.
- Improve the quality of work produced.
- Improve workplace behaviour.
- Make it is more enjoyable for people to work where and how they work.

With the growing sophistication of networked information systems underpinned by the complexities and invisibility of technological processes, focusing on ethical standards of professionalism and the ability to recognise potential ethical issues occurring in both systems and their supporting architectures are of paramount importance. Understanding the impacts of decisions in development, the flows of information and responsibilities of the providers – whether technology developers or infrastructure providers – is key to providing systems that contribute to societal good. Issues of professional responsibility, knowing where – and with whom – accountability rests, and developing systems for transparency and accountability, are the crucial immediate challenges for ethical accountability.

It is commonplace to say that today's information society poses societal, ethical and professional challenges to those working in

the field of computing. IFIP's Technical Committee 9 on ICT and Society has therefore been keen to select three specific challenges, and to explore the issues they face relating to social accountability and professional ethics through the presentation of relevant papers in the context of the ETHICOMP 2015 conference: creating a framework for ethics of computing; cloud computing; and the challenges of virtuality.

6. CONCLUSIONS

There are clearly issues at stake in resolving whether it is at all possible, and how, to arrive at an accepted international code of ethics for the computing profession. Considerable understanding has been enhanced and many advances have been made throughout the past more than two decades. This paper therefore summarises what has been happening in the field of professional ethics – and the codes that reflect those changes – on the international scene for a 25-year period. The principle descriptions include the work of IFIP but also of the ACM, CEPIS, and the IEEE. Suggestions are made on how to move forward, particularly in terms of creating two-part codes of ethics that distinguish between the internationally acceptable and the locally feasible. Three particular fields have been selected as examples for possible on-going activity in the field of ethical dialogue: a framework for ethics of computing; cloud computing; and virtuality. It is our hope that not only the search for an acceptable international code, will make progress through constructive discussion, but moreover that conscientious, applied, examination of key ethical and societal challenges will continue to take place in the future both in practice and on practice.

7. DISCLAIMER

The views developed in this paper are those of the individual authors, and they do not reflect necessarily the position of either IFIP or the other computing associations cited.

8. REFERENCES

[1] IFIP (2015). Update on IFIP Domain Committee on Cloud Computing. IFIP News. June 2015, p. 6. http://www.ifip.org/images/stories/ifip/public/Newsletter/2015to2016/news_jun_2015.pdf.

[2] Anderson, R. (1995). The ACM Code of Ethics: History, Process, and Implications, *Social Issues in Computing*, McGraw Hill, New York, pp. 48-72.

[3] Berleur, J. and d'Udekem-Gevers, M. (1994). *Codes of Ethics, or of Conduct, Within IFIP and in Other Computer Societies*, *13th World Computer Congress 1994*, pp. 340-348.

[4] Berleur, J., Burmeister, O., Duquenoy, P., Gotterbarn, D., Goujon, P., Kaipainen, K., Kimppa, K., Six, B., Weber-Wolff, D., and Whitehouse, D. (2008). *Ethics of Computing Committees: Suggestions for Functions, Form, and Structure.* Publication of the IFIP Task Group on Ethics of Computing and its Special Interest Group IFIP-SIG9.2.2, IFIP Press, Laxenburg, Austria.

[5] Berleur, J., Duquenoy, P., Holvast, J., Jones, M., Kimppa, K., Sizer, R., and Whitehouse, D. (2004). *Criteria and Procedures for Developing Codes of Ethics or of Conduct (To Promote Discussion Inside the IFIP National Societies), On behalf of IFIP-SIG9.2.2*, IFIP Press, Laxenburg, Austria.

[6] Berleur, J., Duquenoy, P., and Whitehouse, D. (1999). *Ethics and the Governance of the Internet. To Promote Discussion Inside the IFIP National Societies*, IFIP Press, Laxenburg, Austria.

[7] Brunnstein, K. and Berleur, J. (eds.) (1996). *Ethics of Computing: Codes, spaces for discussion and law*. Chapman & Hall on behalf of the International Federation of Information Processing (IFIP), London.

[8] Burmeister, O.K. (2013).Achieving the goal of a global computing code of ethics through an international-localisation hybrid. *Ethical Space: The International Journal of Communication Ethics,* 10, 4, 25-32.

[9] Capurro, R. (2008). Intercultural Information Ethics, *The Handbook of Information and Computer Ethics*, eds. Himma, K.E. and Tavani, H. T, Wiley, New Jersey.

[10] GESI (2015), #Smarter2030, ICT Solutions for 21st Century Challenges, http://gesi.org

[11] Gotterbarn, D. (1996).Software Engineering: The New Professionalism, *The Responsible Software Engineer*, ed. Colin Meyer, Springer Verlag.

[12] Holvast, J. (1996). Codes of ethics: discussion paper. In *Ethics of Computing: Codes, spaces for discussion and law*, J. Berleur and K. Brunnstein. eds. Chapman & Hall on behalf of the International Federation of Information Processing (IFIP), London.

[13] IFIP (2011). *IP3 Application and Assessment: a scheme for the Recognition of IT Professionals Guidelines*. Appendix 1 - IP3 Standards Document - Code of Ethics checklist.

[14] McLaughlin, S, Martin, S. et al. (2012).European Commission Report CEPIS, *e-Skills and ICT Professionalism, fostering the ICT Profession in Europe*, Final Report, May 2012 http://www.cepis.org/media/EU_ICT_Professionalism_Project_%20FINAL_REPORT.pdf.

Amazon and the Self

Andrea Resca
LUISS "Guido Carli" University
Viale Romania, 32
00197 ROMA - Italy
aresca@luiss.it

Bendik Bygstad
University of Oslo
Postboks 1080 Blindern
0316 OSLO - Norway
bendikby@ifi.uio.nol

ABSTRACT
In his classic book on the information society Castells described the relationship between"the self and the net". The rise of Internet giants such as Google, Apple and Amazon, have been spectacular successes on an unprecedented scale. However, the relationship we have with these companies is characterized by ambiguity; on the one hand it is part of the daily life, such as searching for information or buying a book, through personalized and easy-to-use services. On the other hand, the dominant position of these actors raises concerns on privacy, and monopoly power. In this paper we try to make sense of these asymmetrical relationships, and propose a way forward to achieve a more balanced relationship. Our empirical framing is an investigation of the evolution of Amazon, through the years 1995-2013. We analyze the company at two levels, focusing at first on the individual interaction and then at the level of infrastructure. For the customers the infrastructure is not visible, although the aggregates of personal and behavioral information are the basis for the rich set of "individualized" interactions between customers and company. The basic asymmetry of the relationship cannot be solved at an individual level. We therefore call for a new kind of institution, and discuss some alternative strategies..

Categories and Subject Descriptors
K.4.4 [**Computer and Society**]: Electronic Commerce - *Distributed commercial transactions, Electronic data interchange (EDI), Intellectual property, Payment schemes*.

General Terms
Management, Economics, Security, Human Factors, Theory, Legal Aspects.

Keywords
Internet companies, Networks, Data Aggregation, Social Costs, Amazon, Governance Systems, Ethical Principles.

1. INTRODUCTION
The relationship between the individual and society is a basic issue in law, political philosophy and sociology. The political philosophers of the Enlightment Age called for a contract between society and the individual, and Locke argued that the experience and identity of the Self is a foundation for society [17]. During the past two centuries this relationship has been framed by the nation state; it is the state that issues legislation and enforces it, it is the state that grants you the right to vote and participate in democratic processes, and it is the (modern) state that supports you with health services and welfare benefits.

Some researchers have pointed out that the contract between the individual and society is threatened by globalization [3,18], because individuals increasingly deal with actors – in particular Internet based businesses and organizations - that are not regulated by the national authorities of the country of the individual. The Internet has greatly increased the area and types of transactions outside the jurisdiction of the nation state.

The philosopher Luciano Floridi used the term infosphere [10] to denote how the growth of ICT has fundamentally changed the global systems, and how the "Westphalian system" of nations has been weakened and led to a "fourth revolution in our self-understanding". He pointed to three aspects of this development [3 p.56]:

- Power: The democratization of information has led to a multitude of non-government agents, which we, as Internet users and participants, increasingly relate to in our daily lives.
- Space: ICTs de-territorialize human experience, making borders porous or irrelevant, and creates more tensions between states and corporations.
- Organization: ICTs fluidify the topology of politics, in the sense that technology aggregate and disaggregate groups around shared interests.

The current discussion on privacy after the Snowdon scandal raises many questions [29]; will everything on the net be used (and misused) by public surveillance agencies and private companies? Are the Internet companies the biggest threat, as Zuboff argues in her description of surveillance capitalism? How can the deeply asymmetrical relationship between the "Big Other" (governments and companies) and the ordinary citizen become subject to democratic and transparent processes, and possible mitigated?

Nowhere is this more visible than our relationship to the Internet giants, such as Amazon, Apple and Google. While our transactions with these actors are convenient and regulated by agreements, the broader relationship is much more problematic. The issue of privacy has been much discussed, but a more general aspect is more important: the relationship is deeply asymmetrical in terms of knowledge and power.

In this paper we investigate this relationship, from an informational and economic point of view. Specifically, to develop our argument we draw on the concept of negative externalities and social costs as representing effects not registered by the market [1]. Transactions with the Internet giants involve factors that spill over the simple delivery of a good in exchange of a specific amount of money.

{Publication}, Month 1–2, 2015, City, State, Country.

2. THE NET AND THE SELF

"Our societies are increasingly structured around the bipolar opposition of the Net and the Self», Castells wrote, in his classic work «The Information Society» [2 p. 3]. The «Net» is the Internet, but it also denotes the institutions and structures of globalized capitalism. The demise of nation states and the rise of multinational corporations make way for a «net-like» economic and cultural structure, electronically mediated by information and communication technologies. Against this stands the Self, which symbolizes the efforts that people make to create their identities in a changing and fluid world, building on more stable attributes, such as sex, religion, place or ethnicity. The Self is not passive or isolated; rather it is a powerful enabler of societal change, and in particular social movements, such as feminism and environmentalism.

Castells dramatically positions these two forces as the key contradiction and tension of the modern world. In this study we conceptualize the relationship in a simple figure, as shown below.

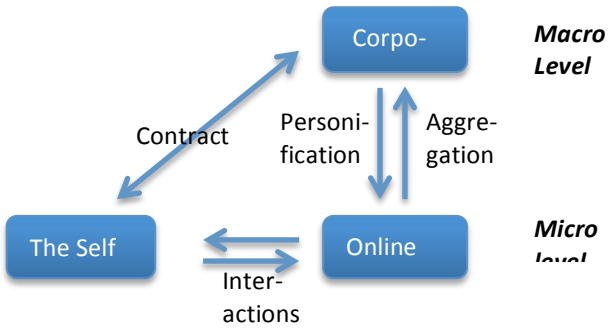


Figure 1. The nature of interaction between Amazon and the Self

We relate to the Internet Corporation at two levels, as illustrated in Figure 1. At the micro level we conduct various transactions; we search for information at Google, we buy books at Amazon, and we download music from iTunes. In this sense it may look similar to shopping grocery or buying a pint at the pub. However, the function of the macro level of the Internet companies is very different: as shown in figure 1 the information of these transactions is accumulated and refined in sophisticated ways, and combined with personal information to create profiles of behavior. This information is used for more personal marketing ("other people who bought this, also bought…."), but it may also be sold to other companies or handed to government authorities, as the Snowdon case showed.

In principle these uses are dealt with in the "contract" with the corporation, that is, the long and legalese agreement that most users don't attempt to read, but just checks OK. In practice the contract is quite asymmetrical; it grants the corporation the right to the information, while the individual user is left with hoping that it will not be misused. For a non-US citizen it is also clear that the reach of the "home" State is limited, to say the least.

At the beginning of the '90, Claudio Ciborra [4] discussed whether an electronic market place should be another public good to be provided by the government or by the invisible hand of the market. In this respect, the stock exchange can be considered a typical example as transactions are concentrated, regulated (e.g. rules against stock manipulation), and transparent. No doubt, the invisible hand has prevailed and transactions are fragmented into a large number of electronic market places rather than concentrated into few ones.

Amazon is one of the main players in this respect and the economic perspective can be of some help for investigating the characteristics of the interaction between users and electronic markets. Specifically, the economics of externalities, as the discipline that studies the side-effects of the market leading to social costs, is considered helpful for this proposal. Social costs exist simply because the market is unable to register some phenomena [24]. Pollution is a typical example for differentiating private costs and social costs as the cost of a specific production is not only related to what is paid to realize it (private costs) but also what has to be paid by the society at large in terms of the recreational usage of a specific area or a damage to another industry like tourism (external costs) ([8]. Another example is related to child labor. Differently from the previous case in which is difficult to assign a price to pollution, here market mechanisms operate but this situation clashes with the acknowledged right of children not to work. The market leads to social costs that are inconsistent with social values [22]. This means that some transactions may be forbidden, some property rights may be denied and other rights may be acknowledged. Therefore, there is no market as such but different possible markets each one with its legal-economic nexus [24]. At this point two main factors determine social costs: the imperfections within the market as in the case of pollution and the legal-economic nexus of the market that may not adequately reflect social values as in the case of child labor [24].

According to Ramazzotti et al. [23] social costs have four common features: 1) they affect a great number of people; 2) they do not only relate to allocative efficiency but also the terms of distribution, employment and the stability of the economy; 3) the collective decision-making is affected negatively and some sections of the economy gain from the existence of social costs; 4) social costs tend to feed back on the economy reinforcing their negative effects.

3. TECHNOLOGICAL INNOVATION AND IMPERFECTIONS WITHIN THE MARKET

Any introduction of a technological change tends to carry externalities and social costs, in the sense that an innovation may affect in an uncompensated way agents who have been involved in this new setting [1,4]. Typical questions related to externalities are: who is liable and should pay for the damage? What is the best solution to reduce the damage? How, and through which organizational arrangements can affected individuals react? Ramazzotti [5] emphasizes the importance to consider externalities and how the social arrangement selected assigns them to the parties involved.

Essentially, the solution is seen in the emergence of property rights through social arrangements that enables a bargaining process that, in turn, makes a cost explicit. The so-called internalization of social costs represents the possibility to convert costs that can be fixed with difficulty into costs that are measurable and allocable. On the other hand, as already mentioned, organizations, both private and public, tend not to reduce externalities for this reason. Usually, they are quantified only in consequence of long litigations and it is the judge rather than market mechanisms that define them. To change the situation is necessary that individuals bearing externalities voice their interests or the presence of environmental factors such as a new legislation and the exercise of collective action. The question is

even more relevant when externalities and related social costs can be detected with difficulty as in the case of innovation driven by IT. They are often subtle and to inform about them is problematic as in the case of the role of metadata generated by a electronic transactions. Extra-market information (e.g. metadata) or information about the market is just as important as market information (e.g. price and quality of the goods) or information from the market [24]. IT users could not be aware about the implications related to the use of a specific application or to evaluate, discuss, and communicate them in an appropriate way. Technological progress in itself is an additional element to be considered for the recognition of social costs. The adoption of new software and hardware requires a continuous learning and engagement.

Even in the case in which externalities are disclosed easily, an additional problem for establishing a suitable bargaining process is related to the need for collective action [20,21]. Social costs tend to be diffused and IT users to act autonomously so that becomes challenging to coordinate their voice. Besides, collective action is subject to the "free-riding" phenomenon in which some of the users do not bear the cost of trying to change the situation [26,27].

To sum up, in presence of externalities and then social costs, three situations could emerge [4,7]: 1) "no formal bargaining"; 2) "explicit contracts"; 3) "state regulation". In the case of "no formal bargaining", the existence of externalities due to the use of IT is not acknowledged. Users are required to adapt following technological requirements. A form of resistance is represented by a partial and watchful use of the system or by individual bargains regulated by ad hoc contracts. The agreement upon "explicit contracts" requires the presence of a representative organization, like unions in the company that are able to actually represent users in the bargaining process with management. The objective of the bargain should be to re-arrange properties rights emerged in consequence of the IT use. In this respect, what is fundamental is an investigation of the system while it functions. In the case where organizations have no incentives to reduce externalities borne by IT users, regulation represents a solution forcing them to modify terms of use. According to Ciborra [4], in two specific cases "state regulation" is considered essential: in case it is impossible to exclude anyone from the impact of the technology (e.g. citizens' privacy); in case the gap in resources and information between the organization and users prevents to initiate and maintain a coherent collective action by citizens to influence the IT use provision.

Amazon's evolution confirms that, substantially, the no formal bargaining modality has imposed. The presence of externalities and social costs shouldered by users interacting with this Internet giant has not been recognized so important and a legal disclaimer regulates them. The point, now, is to investigate how Amazon operates focusing on its governance system and particularly focusing on its meta-governance or values, norms, and principles at the basis of the governing approach. As our analytical lens, nine principles are considered [16]. Three principles (transparency, efficiency, accountability) concern governance elements that are represented by images or ends and goals of governance, by instruments or tools and solutions at disposal to achieve them, and action or the putting of instruments into effects through policies or the mobilization of actors in new directions. Three principles (respect, equity, inclusion) concern governance modes or, in other words, how interactions are institutionalized and in which type of institutions: hierarchical governance (coordination is hierarchical and formal); self-governance (horizontal and self-regulating coordination); co-governance (coordination is horizontal and carried out through mutual adjustments). Three principles (effectiveness, responsiveness, moral responsibility) concern governing orders or levels of activities for solving problems or taking opportunities. The first level (first-order governing) relates to day-to-day affairs, the second (second-order governing) questions the institutional setting at stake, and the third (third-order governing) deals with the entire governance exercise and the application of governance principles.

4. RESEARCH STRATEGY

In order to understand the nature of the relationship between the self and the net (micro level), we investigate the interaction processes between users and one of the most successful dotcom companies. However, the aim is also to describe Amazon's evolution on the basis of these processes from 1995, the foundation year, to 2013 (macro level). Several reasons are at the basis of this in depth case study [28]. It is a critical case study. The extensive use of solutions for users' profiling and the size achieved at a global level in a limited span of time suggest the relevance of the case. Amazon is also a representative or typical case study as companies such as Google, Facebook, and Apple do not differentiate significantly as far as interactions with users are concerned. It is also a revelatory case representing one of the first examples in which an online bookstore has transformed in something more. The idea is also to build a longitudinal case. At first the governance system that have characterized Amazon so far is introduced and then is proposed an hypothetical governance system in order to deal with the asymmetrical nature of the relationship with users leading to an "exploratory" study. Due to the fact that the present case involves more than one unit of analysis, it can be included among embedded cases [28]. What are at stake is both Amazon's interactions with users (micro level) and their effects on its evolution (macro level).

With all this considered, the research question relates to the possibility to figure out a governance system able to keep under control externalities and social costs caused by the asymmetrical relationships in which users are involved and then monopolistic positions of Internet giants. In this respect, two main propositions are posed. The first one concerns a governance system proposal able to modify the nature of interactions between users and Amazon. The second one, related to the former, concerns the possibility to mitigate the monopolistic positions that Internet giants are acquiring in these days.

The theoretical framework in order to deal with these propositions is based on externalities and social costs. Due to the fact that they throw light not only on imperfections within the market but also on the legal-economic nexus at its basis, a comprehensive interpretation of the phenomenon in question emerges. However, this framework falls short to actually propose a solution for keeping under control externalities and social costs other than suggesting a list of possible situations. This is the reason why a governance theory is introduced. A further aim is to see whether the theoretical framework proposed including the governance system can be generalized to other Internet giants such as Google, Facebook, or Apple.

In order to analyse the macro level, Amazon's evolution is taken into consideration on the basis of the concept of platform and the concept of infrastructure. They are used as metaphors for investigating the relationship between technology and the business environment. At the basis of the concept of platform and infrastructure there is the concept of IT as the elements of the organized activity. Routines, structures, processes and transactions are examples in this respect. Differently, an

application represents an organizational function (a set of routines, structures, processes and transactions) such as marketing, finance, production, sales, etc. [14]. We should expect that a platform would represent an established set of functions that interacts with the environment. Rather, this metaphor suggests an alternative modality in the configuration of functions. A platform is considered a sort of springboard: "oil platforms, platforms at bus and railroad stations, platforms for launching missiles, etc. as a basis to stand on, perform actions on top of, or be used to enter another "domain" [13]. In this way, organizational functions are conceived as the footing to deal with the business environment rather than a static solution.

Hanseth et al. [13] discuss also the concept of infrastructure. The term infrastructure, normally, suggests the series of services that characterize modern society such as water and electricity supplies, public transportations, road networks etc. At a first look, a significant difference between the concept of platform and the concept of infrastructure does not emerge. However, a platform tends to support a specific endeavor (oil extraction, missile launching, train access etc.). In contrast, an infrastructure tends to support a large community or a society. It constitutes an underlying level for supporting human activities in a larger social order. Infrastructures can also work as a platform if they are "built on top of and by combining or integrating existing infrastructures"[13]. In organizational terms, an infrastructure, differently from a platform that outlines the organizational functions/environment relationship, is conceived by the configuration of the entire system in which the organization operates.

The fact that the proposed governance system envisages the reduction of externalities and social costs contributes to the validation of the study. In other words, it is possible, in some sense, to expect that the solution proposed is significant in this respect. Further, the validity of the case is demonstrated by the possibility to generalize it to the domain of the Internet giants. Finally, the operation of the study can be repeated as this is essentially based on a document analysis and our primary source was the Internet. Since its foundation, Amazon has been covered intensively by all media and a large amount of data and information is available. Specifically, three main sources have been used in evidence collection: Wikipedia, Amazon's press releases, and Amazon's balance sheets and related documents. All main Amazon's applications, solutions, partnerships, etc. have an entry in Wikipedia allowing a cross-search. Being listed in the stock exchange, Amazon is bound to inform shareholders about the financial situation and press releases are continuously issued for updating on current activities. Finally, the "Wayback Machine" application (by the way, provided by an Amazon's company) has been used in order to compare Amazon's web pages in the course of the years. However, a document analysis has limits. First hand data and information were not available even though the intention was to rely upon accredited sources such as Wikipedia and balance sheets.

5. THE AMAZON'S MICRO AND MACRO LEVEL ANALYSIS

The analysis of the relationship between the self and the net (e.g. Amazon) (micro level), and the evolution of this internet giant (macro level) according to the platform metaphor and the infrastructure metaphor is the way to throw light on the nature of externalities and social costs.

The platform metaphor and the infrastructure metaphor constitute the two main concepts for investigating Amazon's evolution according to three periods of time. Specifically, the question is whether this study succeeds to give a reason for the several steps that have led the transformation of an online bookstore into a provider of a large range of online but also offline services.

5.1 1994 – 2001

This period is characterized by the crucial role played by the e-commerce website for selling books, at first, and then also DVDs, electronics, toys and games. Amazon is a typical dot com company and among the most famous. Founded in 1994, it realized that the development of the internet could be a new opportunity for envisaging new business models such as the possibility to sell books according to novel modes. Therefore, the e-commerce website represents the platform as intended by Hanseth et al. It is the basis on which organizational functions were arranged in order to deal with the environment. The environment, at first, was represented mainly by internet users as potential customers and then also by merchants (ZShops), by customers (Amazon Marketplace for selling used and collectible items), and by partners (members of the Amazon Associate program).

The infrastructure metaphor [13] suggests the importance to support a large community or a society. Due to Amazon's website, customers are not only the addressees of final products but also sellers of used and collectible items; traditional book sellers are not only Amazon's competitors but also contribute to offer a larger selection of titles to readers; website owners are not only clouded by Amazon but also have the possibility to act as a mediator with internet users (members of Amazon Associate). Amazon website can be considered an infrastructure for the mobilization of a large number of actors. Users have different roles according to the situation. An Amazon's customer can transform into a supplier and a competitor into a partner. It is in this picture that the characteristics of the relationship between the self and the net was taking place. Specifically, the series of mergers and acquisitions realized during this period such as that one of Sage Enterprise, Drugstore.com, Alexa Internet, and Leap Technology were focused on the possibility to trace traffic patterns of Amazon's website user community. The aim was to aggregate technological features and software solutions able to profile users for promoting new possibility of transactions. The conception of user not only as customers but also as competitors and partners has favored this process and what can be emphasized is the importance of aggregating a large number of users with different roles but part of a same endeavor toward business development in scope and scale.

5.2 2002 – 2006

The e-commerce website has continued to evolve also in this period due to new categories of products on sale (apparel & accessories, sports & outdoors, jewellery, etc.) and new features. The "Look inside the book" feature advanced to the "Search inside the book" one, Amazon Prime membership program was introduced as well as Amazon ProductWiki and Amazon Connect to scratch shipping-costs for a flat fee and involve internet users in the evaluation of products on sale. The "Look inside the book" and then the "Search inside the book" have inaugurated the digital content creation and with Amazon Unbox digital contents such as music and videos are available on the e-commerce website.

Two main factors suggest that new ways to deal with the environment have emerged: mergers and acquisitions of companies in the on-demand content (books, videos, and CDs)

business and the making available of Amazon e-commerce applications to developers and website owners. It is provided the opportunity to customers not only to buy and sell contents but also to produce them directly as well as the offer of cloud computing services as a commodity. Now, the e-commerce website has been integrated with a range of applications for offering new modes for the storage, elaboration and management of data and for the production of contents. Due to the new traits of the platform, the reference environment has changed too. Companies in search of novel solutions in the IT management domain and internet users interested in the production of contents also represent Amazon's environment of this period.

The infrastructure metaphor suggests that the construction of communities has carried on. Amazon Product Wiki and Amazon Connect have contributed to involve customers commenting items on sale and promote forums between authors and readers. However, the decision to allow developers and website owners to access Amazon's web applications has attracted a new type of partners such as the IT expert community that has played an important role for approaching customers of cloud computing services. The same role has been played by the Mechanical Turk service. Also this period has seen a development in scale and scope of the self/net relationship. New types of users have been targeted and new technological features have been introduced in order to enrich this relationship improving users' experience and the possibility to figure out new opportunities of interaction and exchange. In fact, Amazon has continued to pursue the strategy to aggregate internet users as customers, sellers, and mediators due to the development of the e-commerce website. This aggregation process has been also characterized by the provision of web applications and the possibility to realize contents on demand to be put on sale, eventually, on the same Amazon's e-commerce website.

5.3 2007 – 20013

Amazon's evolution has continued also in this period. This was due, mainly, to the full operation of retail websites, to the provision of e-commerce and logistics solutions, to the launch of the Kindle series, and to the entry into the publishing sector. Operating retail websites meant that the Amazon e-commerce website and related logistics have been replicated for companies such as Sears Canada and Marks & Spencer. Now Amazon does not only provide a marketplace but also a customized e-commerce site complete of warehousing and shipping management. The creation of digital contents has been integrated with the production of electronic devices for accessing them and the entry in the publishing sector is linked to this strategy too. In this way, the self/net relationship has been extended as well. Customers/users of electronic devices have become new sources for gathering data and information and then able to consolidate already available sources. This trend has also been favored by the fact that Amazon now is also a provider of e-commerce websites taking charge of the logistic side at different levels and a publisher both of online contents and paper-based contents. The community building has continued also in this period even though it has characterized mainly actors already present within Amazon borders. Specifically, the series of acquisitions can be conceived to be led by the reconfiguration of the content business. Amazon is acquiring a relevant position in this market sector being a publisher, a book seller, a provider of services for contents on-demand, and a manufacturer of devices for accessing contents. To sum up, the aggregation process established by Amazon is carrying on redesigning the provision of business services, of purchasing ways, of content production and content access.

6. AMAZON'S GOVERNANCE SYSTEM PRINCIPLES

The question now is to throw light on externalities and social costs that emerge interacting with an Internet giant such as Amazon. The nine principles introduced above related to governance elements, governance modes, and governance orders are used for this proposal. Therefore, at first, the question is whether principles such as *transparency*, *efficiency,* and *accountability* have governed Amazon's evolution as described in the previous sections. According to [16], the principle of transparency should guide ends and goals of governance. The situation in which much is known by many [9] does not seem to characterize the nature of interaction with Amazon as a large part of terms are not disclosed or can be disclosed with difficulty. The conception of users as providers of data and information on which to develop business activities does not emerge with clarity. The principle of efficiency distinguishes the choice and application of instruments and it can be declined according to: cost efficiency; productivity or economic efficiency; and response time or operational efficiency [16]. Probably Amazon is one of the more significant examples in which all these kinds of efficiency have been achieved. Accountability as a principle for governing actions or how instruments are put into effect in order to pursue ends and goals can be subdivided into three targets: giving account, holding account, and direction of accountability. The giving account about the self/net relationship in the case of Amazon is not exercised and actions to hold it to account are limited to station-led policies due to the privacy issue rather than citizen-led activism.

Principles at the basis of the governance modes or structures within which actors operate emphasize the role of respect, equity, and inclusion. Respect should typify self-governance as the structure in which the self/net relationship takes place. Respect is seen as the consideration for or avoiding intruding on persons and things. It goes without saying that this principle is not followed properly in the Amazon's case where technological features sneak in during interactions with users. Equity can be analyzed according to procedural equity and outcome equity. The former deals with procedures and the latter with criteria in the distribution of costs, benefits, hardships and burden sharing [2]. It is questionable if the procedure at the basis of the self/net relationship is equal other than the sharing of costs and benefits considering benefits for are tangible differently from costs. Inclusion is recognized a principle for balancing power relations and the fact to have a voice about issues and decisions with which one is concerned [6]. This is typical of the co-governance mode in which members are actually involved in the decision-making processes. In contrast, Amazon, as a typical corporation, substantially does not contemplate inclusion or can be a consequence of an act of magnanimity.

Effectiveness, responsiveness, and moral responsibility are the principles that should guide the levels of activities. Effectiveness relates to the activities for problem solving and opportunity creation. In this respect, Amazon has succeeded to provide to users a large range of business opportunities other than to envisage innovative ways to experience consumption. As mentioned in the section above, entire business sectors have been reconfigured as well as distribution channels. Responsiveness is the principle that represents the capacity of an institution to respond to wishes of the governed and, at the same time, to stimulate the governed to measure taken by their governors. Also in this instance, Amazon has been an exemplary case of both aspects of responsiveness. Moral responsibility is conceived as the

principle that should evaluate the entire governance exercise. An exercise that should be ethical and be justified according to generally accepted values. Can we say that Amazon is responsible from a moral point of view? It's hard to say even though doubts have emerged in the study of the case suggesting that other principles prevail as demonstrated by the presence of externalities and social costs raised from this analysis.

7. TOWARD A PERSONAL DATA BANK

At present, digital personal data generated when we make a purchase, visit a website or use a social network are not under our control. Due to the acceptance of a disclaimer, these data are in the hands of retailers, governments, insurance companies etc. Turning to an analogy with financial assets, the present situation sees corporations and governments managing assets (digital personal data) of Internet users to their advantage. In contrast, in a capitalist economy, citizens are in control of their financial assets and decide how to spend or invest their money. Financial institutions emerged in order to secure financial resources and mediate between how as resources available (money savers) and who is in need (entrepreneurs). At this point, the aim is to figure out a similar institutional context for managing digital personal data [12]. Citizens should be considered the only subjects who have the right to collect and integrate all their personal information (retail purchases, phone records, smartphone data, medical records etc.) However, there are two main differences between personal financial assets and personal data assets. The first one is that information related to financial assets is under the control of financial institutions, subject to strict privacy rules and at disposal of money savers. In contrast, digital personal data acquired by Internet companies are not at disposal of Internet users. According to the right of digital self-determination users should be entitled to have a copy of these data. Probably, a legislation is necessary to obtain these copies as companies should share data at the basis of marketing polices and better customer services. However, the question of transparency is gaining momentum. Further, this change is favoured by the Vendor Relationship Management (VRM) or solutions based on software tools that aim to provide customers with both independence from vendors and better means for engaging with them [25]. Copies of digital personal data are resources that have economic value and are equally distributed. Therefore, an opportunity for instantiating the political philosopher John Rawls' idea of property-owning democracy [19] is at hand due to the possibility to employ them according to personal preferences and needs.

The notion of property-owning democracy suggests also the corporate governance that the envisaged personal data bank should acquire. The form of cooperative, in fact, promotes citizens' empowerment considering that according to the International Alliance of Cooperatives "a co-operative is an autonomous association of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned and democratically-controlled enterprise. Co-operatives are based on the values of self-help, self-responsibility, democracy, equality, equity and solidarity. In the tradition of their founders, co-operative members believe in the ethical values of honesty, openness, social responsibility and caring for others. The co-operative principles are guidelines by which co-operatives put their values into practice." It goes without saying that the adoption of the cooperative form could represent a further step out of digital dependency that citizens and society have ended up. In this way, individuals have the possibility to legally own or have legal rights of access and control on their data, and be member of a democratic body that has the potentiality for value creation in a context in which legal and ethical norms are exercised as well as privacy.

Considering the role that financial markets have acquired in these days as the fuel for investments and entrepreneurship, it is possible to figure out the possibilities of a personal data bank. This bank could become appealing for several reasons other than dealing with the asymmetry of the self/net relationship. In fact, rather than to be in the hand of few Internet giants, personal data can become available to any private or public actor according to market mechanisms. The potentiality of this environment, as the open data phenomenon suggests, for promoting innovation both in the public sphere and in the private sphere is hard to imagine.

8. THE GOVERNANCE PRINCIPLES OF THE ENVISAGED PERSONAL DATA BANK

The nature of the self/net relationship changes considerably due to the introduction of the personal data bank. Externalities and social costs are kept under control and the analysis of the governance system according to the nine meta-governance principles contributes to this account. Considering, at first, principles related to governance elements or images, instruments, and actions, such as transparency, it emerges that this principle is supported in the case of the personal data bank. The idea is to provide the opportunity to track who, when, and preferably even for what purpose data were accessed and to monitor by any subject his/her personal data. This is a context in which much is known by many [9] differently from the Amazon's case. The principle related to the choice and application of instruments is efficiency. The personal data bank is only an idea, a project, and at a so early stage it is not very significant to take into account this principle. It is possible to state that the principle of accountability or giving account, holding account, and the direction of accountability governs bank actions or the putting into effects of instruments for following ends and goals. The giving account is intrinsic to an organizational form such as a cooperative as well as the possibility to hold it to account and the citizen-led actions rather than the state-led ones.

Turning to the principles at the basis of governance modes or organizational structures within which actors operate, the emphasis is posed on respect, equity, and inclusion. Being an associate to a cooperative means that you are taken into consideration and intrusions on persons and things are limited or object of discussion and then shared according to norms. In fact, associates play a dual role: they are at the same time providers and managers of digital personal data. Procedural equity and outcome equity represent the principle of equity and they are followed as well. We can expect that data of any single subject are dealt with equally as well as the distribution of costs, benefits, hardships and burden sharing that distinguish outcome equity. Inclusion is also promoted. The one head one vote rule is at the basis of the cooperative governance system in which any member has the same right to voice or to act about issues and decisions of his/her concern.

Governance orders or the levels of activities that characterize a governance system should be guided by principles such as effectiveness, responsiveness, and moral responsibility. Can we expect that the personal data bank will be effective for problem solving and opportunity creation? It could provide an environment in which issues raised by the self/net relationship are managed. As far as the question of opportunity creation is concerned, the possibility to offer digital personal data both to private and public

players could open new scenarios for product and service innovation. At this stage, it is difficult to say whether the governance system figured out can be considered responsive. For sure, it is an instance for responding to issues of the governed here represented by the cooperative associates. We can expect also that they will be stimulated by measures taken by governors as involved, even though indirectly, in the cooperative management. Finally, moral responsibility is taken into account. As the general principles that should distinguish the entire governance system, what has been prefigured for dealing with externalization and social costs of the present self/net relationship is substantially ethical and can be justified according to generally accepted values. Recalling the nine principles for meta-governance, a significant difference emerges in comparison with the Amazon's case.

9. CONCLUSIONS

Probably, the governance solution proposed deals with difficulty the self-propelling growth of information typical of the Internet giants[15]. To have at disposal digital personal data in order to create users' profiles for marketing polices and better customer services could be no more useful at this point. Nevertheless, issues raised by the theory of externality and social costs have been managed. What has been described is a shift from a "no formal bargaining situation" to an "explicit contracts" situation. Internet players could be interested in the quality of data available in the imagined personal data bank and then willing to buy them.

The proposed governance solution would allow the so-called internalization of social costs as property rights related to digital personal data would come out. Further, to turn to the organizational form of the cooperative means to promote a legal-economic nexus that protects those social values that are substantially neglected at the moment. Finally, it would be naïve to believe that this solution represents a shelter from the monopolistic positions of an Internet giant such as Amazon. However, a new market, the market of digital personal data, has been outlined and market mechanisms rather than the feudalism of the Internet giants will govern their allocation.

10. REFERENCES

[1] Arrow, K.J.*The limits of organization.* Norton, New York, 1974.

[2] Banuri, T., Goran-Maler, K., and Grubb, M.Equity and social considerations. In J.P. Bruce, H. Yi, E.F. Haites and Intergovernmental Panel on Climate Change, eds., *Climate change 1995: economic and social dimensions of climate change*. Published for the Intergovernmental Panel on Climate Change [by] Cambridge University Press, Cambridge [England] ; New York, 1996, 448.

[3] Castells, M.*The Rise of the Network Society.* Blackwell Publishers, Malden, Mass, 1996.

[4] Ciborra, C.*Teams, markets and systems: business innovation and information technology.* Cambridge University Press, Cambridge, 1993.

[5] Coase, R.H.Problem of Social Cost, The. *Journal of Law & Economics 3*, (1960), 1.

[6] Dahl, R.A.*Democracy and its critics.* Yale University Press, New Haven, 1989.

[7] Evans, J.*Negotiating Technological Change.* European Trade Union Research Institute, Brussels, 1982.

[8] Field, B.C.*Environmental economics: an introduction.* McGraw-Hill, New York, 1997.

[9] Finkelstein, N.Introduction: transparency in public policy. In N.D. Finkelstein, ed., *Transparency in public policy: Great Britain and the United States*. Macmillan Press Ltd. ; St. Martin's Press, Houndsmills : New York, 2000, 184.

[10] Floridi, L.On the intrinsic value of information objects and the infosphere. *Ethics and information technology 4*, 4 (2002), 287–304.

[11] Floridi, L.Hyperhistory and the Philosophy of Information Policies. In L. Floridi, ed., *The Onlife Manifesto*. Springer, Heidelberg, 2014, 51–64.

[12] Hafen, E., Kossmann, D., and Brand, A.Health data cooperatives - citizen empowerment. *Methods of Information in Medicine 53*, 2 (2014), 82–86.

[13] Hanseth, O., Bygstad, B., and Johannesen, L.K.*Towards a theory of generative architecture. A longitudinal study of eHealth infrastructures in Norway.* Oslo.

[14] Hanseth, O. and Lyytinen, K.Design theory for dynamic complexity in information infrastructures: the case of building internet. *Journal of Information Technology 25*, 1 (2010), 1–19.

[15] Kallinikos, J.*The Consequences of Information: Institutional Implications of Technological Change.* Edward Elgar Publishing, Cheltenham, 2006.

[16] Kooiman, J. and Jentoft, S.Meta-governance: values, norms and principles, and the making of hard chices. *Public administration 87*, 4 (2009), 818–836.

[17] Locke, J.*An essay concerning human understanding, 1690.* Scolar Press, Menston, 1970.

[18] Nagel, T.The problem of global justice. *Philosophy & public affairs 33*, 2 (2005), 113–147.

[19] O'Neill, M. and Williamson, T.Introduction. In M. O'Neill and T. Williamson, eds., *Property-owning democracy: Rawls and beyond*. Wiley-Blackwell, Malden, MA, 2012, 320.

[20] Olson, M.*The logic of collective action: public goods and the theory of groups.* Harvard University Press, Cambridge, 1971.

[21] Ostrom, E.*Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, 1990.

[22] Passas, N. and Goodwin, N.Introduction: a crime by any other name. In N. Passas and N.R. Goodwin, eds., *It's legal but it ain't right: harmful social consequences of legal industries*. University of Michigan Press, Ann Arbor, 2004, 279.

[23] Ramazzotti, P., Frigato, P., and Elsner, W.Social costs today: institutional analyses of the present crises – an introduction. In P. Ramazzotti, P. Frigato and W. Elsner, eds., *Social costs today: institutional analyses of the present crises*. Routledge, London ; New York, 2012, 300.

[24] Ramazzotti, P.Social costs and normative economics. In P. Ramazzotti, P. Frigato and W. Elsner, eds., *Social costs today: institutional analyses of the present crises*. Routledge, London ; New York, 2012, 300.

[25] Searls, D.*The intention economy: when customers take charge.* Harvard Business Review Press, Boston, Mass, 2012.

[26] Williamson, O.E.*Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization.* Free Press, New York, 1975.

[27] Williamson, O.E.*The economic institutions of capitalism : firms, markets, relational contracting /.* Free Press, New York, 1985.

[28] Yin, R.K.*Case Study Research: Design and Methods.* Sage Publications, Thousand Oaks, 2009.

[29] Zuboff, S.Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology 30*, 1 (2015), 75–89.

The significance of ICT in the generation of code of conduct: from the perspective of polarization of ICT and organizational citizenship behavior

Sachiko Yanagihara
University of Toyama
3190 Gofuku, Toyama-shi
Toyama, Japan
+81 76-445-6507
sachiko@eco.u-toyama.ac.jp

Hiroshi Koga
Kansai University
2-1, Ryozenji-cho, Takatsuki-shi
Osaka, Japan
+81 72-699-5151
koga@res.kutc.kansai-u.ac.jp

ABSTRACT
The purpose of this study is considering the creation of the Code of Conduct by using ICT. In particular, we consider the ethics of ICT by a case study of Japanese NPO which does a support project of "hikikomori." On a global scale, social withdrawal and reclusion are real problems and in the information society of today, these problems are often associated with modern lifestyle and use of technology. However, in this paper, it is not to discuss the therapeutic point of view of withdrawal. Rather, it is discussed online communication in withdrawal support groups. It is the object of this paper to clarify the process of the Code of Conduct through online communication is formed.

Categories and Subject Descriptors
K.4.3 [**Organizational Impacts**]: Computer-supported collaborative work and Employment

General Terms
Human Factors

Keywords
Code-of-Conduct, Organizational Citizenship Behavior.

1. INTRODUCTION
In the area of information resources management research, information and information systems are often considered separately [1]. Taking this into account, information ethics problems can be broadly divided into the artificial aspects and informational aspects of information systems (IS). With regard to the former, one approach is to equip information systems with an ethical function so that technology systems, which are artifacts, have a political characteristic [2] [3]. The appearance of new artifacts such as surveillance cameras and elaborate humanoid robots (androids) has given rise to entirely new ethical issues. Meanwhile, with regard to the informational aspects, in so far as the digital information provided by IS induces new action, it has the function of promoting organizational practice. Zuboff [4] referred to this kind of function as "informate." Here, digital information can be considered to include ethical issues if we assume that there is embedded normativity in organizational practice induced by displayed digital information.

The purpose of this paper is to examine online communities and illuminate the ethicality of their artificial and informational aspects (embedded normativity in organizational implementation induced by the respective aspects), through a case study. In this respect, this paper can be considered as a case study based on a disclosive ethics approach.

There are two main objectives in this paper. The first is to illuminate the process by which behavioral norms are rewritten as the meaning given to the technological entity of the online community changes and new behavior is induced. The second is to illuminate the ethicality of the informational aspect by considering the characteristics of the background behavioral norms behind natural statements in online communication from the perspectives of morality and spontaneity.

Morality in this case is an element of the exective and/or leadership function [5]. Spontaneity is an attribute not only of leaders but also of the constituent members of the organization [6]. For this reason, this paper will attempt to approach ethical issues by examining the process by which social construction of behavioral norms embedded in organizational practice through information systems occurs, and by shedding light on it from the two perspectives of the lerdership based morality and the constituent members' spontaneity.

This paper will focus on an extremely specialized case: that of the online community of an Osaka-based NPO (Non Profit Organization) that works to support people suffering from *hikikomori*—social withdrawal. This NPO is a specialized case in that it enables hikikomori people to engage in discussion through the online community and form their own sense of place. From there, they can find employment through telework and assist other people suffering from hikikomori in finding employment.

However, the intention of this paper is not to provide an overview of these activities, but to focus on the online community of this NPO as a key factor for enabling employment through telework and to discuss "moral creativeness or creative morality" and "organizational citizenship behavior (OCB)" with a view to

{Publication}, Month 1–2, 2015, City, State, Country.

illuminating the two ethical issues of the artificial and informational aspects of online communities.

2. PRIOR RESERCH

Different organizations can implement the same IS with dramatically different methods of use and effects [7]. For example, email systems enable people to send their opinions directly to the president, and they can also be used to monitor the content of employees' communications. It is easy to conceive that organizations may implement the same kind of IS with different methods of use depending on their different organizational objectives, such as streamlining communication, reducing costs, or consolidating authority [8].

It has been observed that the background reason for these differences in practice is the complex interaction of the organization's values being embedded in the IS and new values and practices being created by the content generated by the IS [9]. Put another way, the discourse in which the organization regulates the IS and the IS transforms the organization is based on a perspective that recognizes both the IS and the organization as separate entities and seeks to illuminate their mutual interaction [10].

However, recently a new perspective has emerged, in which the IS and the organization are treated not as separate entities, but as a composite that forms a complete whole [11]. In other words, it is a perspective, based on Actor Network Theory, that understands that "utilization of an IS, that is, the practical implementation of an IS, is produced performatively."

From this perspective, an IS has no meaning until it is introduced to an organization. Alternatively, directly prior to its introduction to an organization, the IS itself does not have an essential meaning or character, but rather through its practical implementation, the meaning and essence of the IS are created (recursively) within the organization. It is a research perspective focused on so-called social materiality.

Incidentally, in the discussion of social materiality, the discussion on behavioral norms and information ethics is insufficient. Therefore, the endeavor of this paper can be considered as a groundbreaking research for incorporating an information ethics perspective into the discussion of social materiality.

There are several key concepts for considering information ethics in organizational practice using an IS. In this paper, however, we will make use of two key concepts from the field of organization theory research.

The first concept is that of moral creativeness or creative morality, from the founder of modern organizational theory, C.I. Barnard [5]. The individuals who make up an organization have relatively stable and fixed values. Barnard refers to these as morals. Then, the function of leadership is to integrate the wills of the various individuals into the direction of the organization goals, and drive them forward. This kind of leadership function that gathers individual wills and focuses them in a single direction is called moral creativeness or creative morality. The model values imparted to the organization by its leaders are precipitated in the organization through its daily organizational behavior, and form the base of organizational activity. As a result, the organization has transformsd to the institutionalization [12]. Therefore, behavioral norms and information ethics related to use of IS are formed and fixed through organizational practice, including leadership.

The second key word is OCB, which is introduced by Organ [6]. Organ [13, p.95] has defined OCB as "individual behavior that is discretionary, not directly or explicitly recognized by the formal reward system, and that in the aggregate promotes the effective functioning of organization" and "performance that supports the social and psychological environment in which task performance takes place."

His followers also went on to classify the constituent elements of OCB as follows: 1) Helping: Altruistic behavior of voluntarily supporting others who are dealing with problems in work or the organization, 2) Sportsmanship: Behavior of accepting work under impossible or detrimental conditions without complaint, 3) Civic value: Contributing actively and constructively to organizational and political systems, 4) Organizational loyalty: Behavior of spontaneously talking about the good aspects of the company and so forth, 5) Organizational compliance: Accepting rules and procedures and adhering to them sincerely, 6) Individual initiative: Striving to undertake missions outside one's own duties while maintaining one's enthusiasm and effort in conducting one's own duties, and promoting the same actions to others, 7) Self-development: Voluntarily acting to improve one's knowledge, skills, and abilities [14].

3. CASE ANALYSIS

3.1 Case Background

This paper takes the approach of a case study. However, it is an extremely special case: that of the online community in an Osaka-based NPO in Japan for assisting *hikikomori*—social withdrawal.

Hikikomori was once regarded as an interesting phenomenon unique to Japan. Today, however, the same phenomenon is said to be observed in other countries too [15]. Recent studies have shown that hikikomori is a variety of onset social anxiety disorder, and is therefore not a special phenomenon [16].

The Japanese Ministry of Health, Labour and Welfare published a guideline in 2010 to reflect these results. The guideline defines hikikomori as follows (Translated by author).

A concept of a phenomena indicating an avoidance of social participation (e.g., schooling, including compulsory education, employment, including temporary employment, and interaction outside the home) as a result of various factors, where, in principle the person has stayed inside his or her home for a period of six months or more. Hikikomori is considered to be a non-mental illness phenomenon, distinct from reclusive symptoms based on positive and negative symptoms of schizophrenia; however it should be noted that in actuality there is significant potential for hikikomori to include pre-diagnosed schizophrenia.

Currently, hikikomori research is conducted not only in the field of clinical medicine, but is also widely studied in fields such as nursing care and social studies. However, there is insufficient data from proper national-scale studies providing epidemiological findings, leading to confusion surrounding the above definition (whether it includes DSM[1] or does not), a need for cohort studies

[1] DSM (The Diagnostic and Statistical Manual of Mental Disorders), which is published by the American Psychiatric

to clarify environmental and genetic causal factors, and a need for experimental planning using random samples [17].

Of course, this paper is not a discussion on the epidemiology or psychology of hikikomori. It is intended to consider from an information ethics perspective the process by which sufferers of hikikomori use social network services (SNS) to start businesses. To illuminate examples of information ethics in organizational practice, we will make use of a case study discussion. However, we also hope that the discussion of this paper will help to support hikikomori sufferers.

In Japan, one approach for helping hikikomori sufferers that is drawing attention is that of creating a place for the sufferers themselves. The case examined in this paper is that of the NPO WISA (Wakamono International Support Association), which would be the unintentional author of this approach. Here, hikikomori sufferers create their own place, find employment through teleworking, and introduce other hikikomori sufferers to telework[2].

However, in this paper, rather than introduce the special activities described above, we look at the significance of the online community in the birth processes of the NPO. In particular, we discuss only the "OCB" and "creative morality" involved in this process.

3.2 CaseOverview

WISA is a registered NPO that works to support young people who are themselves described as hikikomori. As noted previously, hikikomori find it remarkably difficult to communicate with others and since they tend to avoid such interaction, it is difficult for them to hold fulltime jobs. However, WISA focuses on the expert capabilities of hikikomori and runs an operation providing work that can be done at home through online education and communication.

WISA's predecessor was launched in 2009 as the private organization *Sol Life Net*, which was established by calls over the Internet centered on the NPO's director, Mr.Yokoyama, who has experienced hikikomori himself. After graduating from university, he joined a production company. Eventually the long working hours took a toll on his health and he was obliged to take a break from work. Although he returned to work, he was unable to adapt to the atmosphere at the workplace after that, and left. That was the start of his hikikomori lifestyle. He retreated into the world of online games. However, he was not alone. He began chatting with people he met in the online gaming community.

The turning point came from a line from an animated television show. Members of the online community often provide live commentary while watching animated programs or exchange their impressions. The main character in the animation "Eden of the East[3]" which dealt with a group of NEETs[4], said, "There are too few people willing to do something for a loss." Mr. Yokoyama decided to take on such a role himself. He called upon the acquaintances he had met through online gaming through the Game Community (the SNS named *the Otaba* for enthusiasts of gaming)[5]. Twelve hikikomori sufferers and their families came together to create a "community space." This was the birth of Sol Life Net, the predecessor of WISA.

At first, the activities involved planning of town walking events and barbeques to deepen exchanges. This was because fostering relatedness was considered important. Through the activities, Mr.Yokoyama became deeply impressed with the diversity of individual skills, just as with the online gaming community. He sought to find (or if necessary to make) a place where these capabilities could be exercised. This was how he came to attempt to create employment through telework.

Actually, there was a very painful period. His efforts to find work saw him embark on cold-call sales every day. However, once his activities gained recognition, the work steadily increased. In 2010, as the work began to become steady, the group acquired NPO status.

As an NPO, the organization's activities are as follows: 1) supportting for households in poverty and young hikikomori people, 2) creation of IT-enhanced towns, and 3) promoting telework.

Mr.Yokoyama's vision was to draw out the individual capabilities of each hikikomori to help them find work. This activity was highly regarded as an effort to promote telework, and in 2012 the organization received the 12th Telework Promotion Prize (commendation prize: employment finding and creation division) from the Japan Telework Association. In 2012, the organization changed its name to WISA with a view to expanding its activities internationally. Then, in 2013, it was adopted as an Emergency Employment Creation Fund Project of Osaka Prefecture, and received assistance from the government, whereupon it hired two office administrators.

Currently, WISA uses a dedicated SNS to handle receiving and issuing orders for work (The SNS was built in 2013. A freelance registration website and work received via individual introductions are posted on the SNS.) As of March 26, 2014, 73 people are registered on the website. Around 20 of these are active registrants.

The SNS is not just for business use. It also serves as a medium for irregular plastic model creation or design courses and other social events. In addition to online events using video distribution on the SNS, there are also offline events held at the Osaka office and so forth.

Association (APA), offers a common language and standard criteria for the classification of mental disorders.

[2] The description of the following case in next section is based on our interview investigation at February 28, 2014 and March 26, 2014. We used a semi-structured interview. Interviewees are Mr. Takenawa who is a secretariat of WISA and Mr. Yokoyama who is one of the founders of WISA. In addition, we were referring to the fiscal year 2014 business report, which is provided from WISA. And URL of the web site of WISA is as follows. http://wakamono-isa.com/

[3] It is an animation program in Japan that is aired on Fuji TV in the period of April to June 2009. URL of the web site of this program is as follows. http://juiz.jp/special/

[4] NEETs are the abbreviation for "Not in Education, Employment or Training." In Japan, NEETs mean that yong umemployed that are not in school, not doing the housework, and non-labor force of up to 15-34 years old.

[5] Otaba is a so-called geek (Otaku, in Japanese) for SNS. There are three specialty areas: 1) space of interaction, 2) space of discussion, 3) space of creation. URL of the web site of this program is as follows. http://otaba.jp/

Jobs can recruit applicants through this SNS, and people who indicate an interest can receive orders. The content and material for the job can be uploaded on the SNS. People who have received a job are obliged by rules to write a daily report and submit it to the office. However, there are some members who are unable to report properly due to their dislike for communication. Also, for management of administrative operations and meetings, a chat system and email are used in addition to the SNS. Since some members feel pressured in terms of work or psychological factors, they are usually contacted by email. Meanwhile, on the top page of the SNS there is a space for starting a simple conversation, similar to Twitter, where people can easily give a greeting or talk about an interesting event in their day, enabling members to post messages freely.

3.3 OCB in Online Communities

Next, we examine the SNS of WISA from an OCB perspective.

Mr.Yokoyama says that an SNS has key people. The key people are those, for example, who keep track of situations that appear to be growing tense when friction develops between members in their video communication and chats. The characteristics of key people are that they are heavy users of the system, and that they are selfless, says Mr.Yokoyama. Put another way, they are people who can deal with matters flexibly and who naturally draw others together. Even if they do not have strong communication abilities, they have strong business capabilities that enable them to deliver a high quality work performance.

Furthermore, Mr.Yokoyama says that for the SNS to function well, key people alone are not enough. Key people often become involved in disputes on the SNS due to their individuality. At such times, people who can speak or act to help resolve the dispute are needed. He refers to the people who take on such roles as "followers." Followers do not undertake individualistic action on the SNS (e.g., they do not normally make comments). However, they watch over the overall situation and can make comments when appropriate.

The common points between key people and followers are these: first, they are not official duties related to employment; second, they both value the feelings of people involved in a matter; finally, they always have a grasp of the interaction between constituent members in order to make important comments at key junctures. Mr.Yokoyama explains the common elements between key people and followers simply as "the kind of people one hopes to be able to consult with about various things when needed."

Key people and followers make comments that value the feelings of the people involved in a matter so as to prevent them from making the overall atmosphere too threatening. In some cases, they inject a new topic to change the course of a discussion. Mr.Yokoyama describes this kind of function as "self-help ability." We believe this conforms precisely to OCB.

At this point, the source of OCB in WISA appears to lie in experience in online communication. Through repeated daily verbal interactions, sensitivity and behavioral norms are gradually formed. Just as many drops can form a great ocean, or a little income can accumulate into a large sum, through the to and fro of daily communication, sensitivity and behavioral norms are cultivated.

Mr.Yokoyama points to learning through experience in the above-mentioned online community as means for developing self-help ability (that is, the inducement of OCB). He says, "When we follow Internet dependence through to its conclusion, human relationships become deeper". Building human relationships between people suffering from hikikomori is extremely difficult in the real world. Online, however, by continuing to play games, they are able to avoid cutting off human relationships. Even without using language, they are able to continue to construct relatedness through the exchange of items and avatars (characters modeled on oneself). There are 12 members who have been able to deepen their relationships as described above.

Considered in this way, we can see that online communities (SNSs) teach their participants how to behave. It seems appropriate to describe this as the OCB aspect of information ethics. The text and images that flow online (and are displayed on monitors) are more than just data—they carry the thoughts of those who transmit them, as well as convey what they want the recipient to do next (to listen, to help, to praise, etc.); they are "letters of the heart." The ability to respond to them skillfully depends on one's ability to acquire "sensitivity" as one repeatedly receives and answers the letters.

4. Conclusion

In this paper, we have examined the case of the Japanese NPO WISA, which operates a project to support hikikomori through telework. Mr.Yokoyama, who was once a hikikomori himself, was inspired by a trivial event to dedicate himself and start the project. Mr.Yokoyama had some associates that he had met through online games. They were people who had become hikikomori by developing an inability to continue in their social relationships and face-to-face communication. Mr.Yokoyama accepted their current situation as hikikomori and rather than treating them all under the general category of hikikomori, valued relatedness by responding to each of them in accordance with their individual skills. These values resonated with them, and the circle of his associates expanded. In the background to this activity was the artifact of the SNS. The artifact is itself neutral. Mr. Yokoyama imparted new meaning to the artifact when he discovered new values.

Furthermore, the information supplied through the artifact, even if it is as simple as a greeting, "good morning," is an information complex with embedded background information including the flow of communication leading up to that point, the relatedness between the people taking part in the dialogue, the feelings of the sender and of the receiver, and so forth. To explain its meaning and ensure effective functioning of the SNS, the presence of people who undertake OCB (key people and followers) is indispensable. Furthermore, the key to inducing OCB is "a sense of online community." We believe this is born of the amount of experience that members have in online communities. Through experience in using the artifact, users develop the skills of understanding the information embedded in data displayed by the artifact and a sense for responding appropriately. In this sense, we consider that online communities have information ethics.

5. ACKNOWLEDGMENTS

We would like to show our greatest appreciation to Mr.Yokoyama and Mr.Takenawa, at WISA. And this study was supported by JSPS KAKENHI Grant Numbers 26380458 and 26380550, and

by Kansai University's Overseas-Research-Program (April-September /2015).

6. REFERENCES

[1] Synnott, W. R. and Gruber, W. H. 1981. *Information resource management: opportunities and strategies for the 1980s*. John Wiley & Sons, Inc.

[2] Winner, L. 1986, *The Whale and the Reactor: A Search for Limits in an Age of High Technology.* University of Chicago Press.

[3] Akrich, M. 1992. The de-scription of technological objects. In Bijker, W. E. and Law, J. eds. *Shaping Technology /Building Society*. Cambridge: MIT Press, 205-224.

[4] Zuboff, S. 1988. *In the age of the smart machine: The future of work and power*. Basic Books.

[5] Barnard, C.I. 1968. *The functions of the executive*. Harvard University Press.

[6] Organ, D. W. 1988. *Organizational citizenship behavior: The good soldier syndrome*. Lexington Books/DC Heath and Com.

[7] Brynjolfsson, E. and Saunders, A. 2010. *Wired for innovation: how information technology is reshaping the economy*. MIT Press.

[8] Fuller, S. 2012. *Knowledge management foundations*. Routledge.

[9] Orlikowski, W. J. 2009. The sociomateriality of organisational life: considering technology in management research. *Cambridge Journal of Economics*, 34. 125-141.

[10] Leonardi, P. M., Nardi, B. A., and Kallinikos, J. 2012. *Materiality and Organizing: Social interaction in a technological world*. Oxford University Press.

[11] Orlikowski, W. J. and Scott, S. V. 2008. 10 sociomateriality: challenging the separation of technology, work and organization. *The academy of management annals*, 2, 1. 433-474.

[12] Selznick, P. 1967. *Leadership in administration: A sociological interpretation*. University of California Press.

[13] Organ, D. W. 1997. Organizational citizenship behavior: It's construct clean-up time. *Human performance*, 10, 2. 85-97.

[14] Organ, D.W., Podsakoof, P.M. and MacKenzie, S.B. 2006. *Organizational citizenship behavior: its nature, antecedents, and consequences*. Sage. Thousand Oaks, CA.

[15] Kato, T. A., *et.al*. 2012. Does the 'hikikomori'syndrome of social withdrawal exist outside Japan? A preliminary international investigation. *Social psychiatry and psychiatric epidemiology*, 47, 7. 1061-1075.

[16] Ministry of Health, Labour and Welfare. 2010. (in Japanese) *The guidelines for the evaluation and support of withdrawal*. http://www.mhlw.go.jp/stf/houdou/2r98520000006i6f.html

[17] Teo, A. R. 2009. New Form of Social Withdrawal in Japan: A Review of Hikikomori. *International journal of social psychiatry*. 56, 2 (June. 2009), 178-185. DOI=10.1177/0020764008100629

Online Disclosure of Employment Information: Exploring Malaysian Government Employees' Views in Different Contexts

Nurul Amin Badrul
School of Systems Engineering
University of Reading
United Kingdom
n.a.badrul@pgr.reading.ac.uk

Shirley Ann Williams
School of Systems Engineering
University of Reading
United Kingdom
shirley.williams@reading.ac.uk

Karsten Øster Lundqvist
School of Systems Engineering
University of Reading
United Kingdom
k.o.lundqvist@reading.ac.uk

ABSTRACT
This paper investigates the perceptions and behaviour of government employees regarding the disclosure of employment information. Two different contexts, namely 1) official websites and 2) online social networks (OSN: in this case, Facebook) that disclose employees' employment information are selected as contrasting platforms in order to understand how government employees behave towards the same type of information in different contexts. This preliminary study will draw from information boundary theory to discuss how employees' behaviour and perceptions towards a particular attribute vary between the two different contexts. A qualitative strategy was employed and five Malaysian participants from a range of public organizations were interviewed. The results suggest that while all participants were aware of the issue of disclosure, there were mixed responses regarding disclosure. Privacy boundaries were established when employees perceived the context as official and personal. Additionally, participants acted differently when they had the option not to disclose their employment information on their social network accounts. These findings provide knowledge about information disclosure in which privacy implications are influenced by contextual factors.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues – ethics, privacy.

General Terms
Human Factors, Management.

Keywords
Information boundary theory, online disclosure, personal information, online social network, organizational websites.

ETHICOMP'15, September 7–9, 2015, Leicester, UK.

1. INTRODUCTION
The pervasiveness of Internet usage is allowing increasing amounts of individuals' information to be exposed online. Now that the Internet's influence is deeply entrenched in our daily lives, there seems to be a continuous flow of disclosure of individuals' information from various sources. For example, Internet users are required to disclose their personal information in order to complete online transactions with e-commerce firms. Subsequently, with the emergence of online social networks (OSN), the practice of disclosing personal information has become an online culture with global acceptance. OSN users willingly share their personal information with others to fully experience what the OSN has to offer [33].

Amongst the information that is available on the Internet is employment information such as that found in blogs [14] or OSNs [17]. Employment information is any information about an individual's full time job, including (but not limited to) his or her position, job description, salary, organization, working category, grade, staff ID number and working experience. Users of ONS such as Facebook may reveal their employment information on their profile page by providing information about their workplace or professional skills [26].

Since employment information can be seen as an important theme for networking, OSNs that are modelled on exploiting users' employment information are gaining popularity. For example, LinkedIn is a professional networking site that specifically encourages its users to provide employment information in order to fully experience the network [10]. It is the world's largest professional OSN, with over 364 million members in more than 200 countries and territories [20]. LinkedIn users present a professional profile that resembles a CV, with information on current and previous positions. As an OSN that is geared towards professional users, LinkedIn allows its users to engage with other professionals and industry experts to exchange knowledge and ideas.

On the other hand, official websites are another source of information disclosure that reveals employment information. Employment information is often found on public organizational websites. Zhao & Zhao [35] found that information such as employees' full names, job titles and affiliation was available publicly in all of their sample of e-government websites from all 51 states of the United States of America. Although they discovered that in general, the websites have adequate measures for protecting information integrity and privacy, they highlighted several improvements that could be made to discourage potential

threats from intruders and attackers. This paper reports a preliminary study as part of a PhD research project that is investigating the phenomenon of organizational disclosure towards employees' privacy. These findings, however, draw only on employees' behaviour and perceptions of employment information that is revealed on the Internet.

2. ONLINE DISCLOSURE

2.1 Information Disclosure

Researchers have found that online disclosure of personal information can lead to Internet users' privacy concerns [1]. Five dimensions of privacy concerns have been identified: these are concerns about data collection, secondary use, data ownership, accuracy and access [30, 9]. Internet users are reported to be highly concerned about unknown individuals obtaining information about them or their families, threats from hackers accessing their credit card information and information from their online activities [11]. A recent report by Madden & Rainie [22] found that more than 90% of US Internet users are concerned about their information being obtained by unfamiliar parties, which might lead to online privacy risk. Their concern is over the loss of control of their information and the subsequent consequences of its usage. In this digital age, government public records, which were formerly 'hidden', are now readily available and searchable on the Internet. They can be collected by anyone and can be exploited for other purposes. Thus, control of information is the central concept in information privacy concern. In addition, online disclosure may involve risk associated with offline disclosure [13]. This is made possible by exposing geographical information that can facilitate real-world attacks.

Internet users themselves voluntarily post their information on their personal homepages, tweets or social network accounts. Researchers have discovered that the decision to disclose personal information depends on the context or situation [16] and on the expected benefits of the disclosure [3]. John et al. [16] discovered that individuals are willing to divulge information when asked indirectly and when contextual factors (e.g. web interface) are employed. They discovered that sites that give a professional impression garner higher privacy concerns compared to sites with unofficial impression since the formality reminds users to privacy. This suggests that users' information disclosure is dependent on context.

It has been shown that users categorized different types of information to have different potential risks [24]. In doing so, users then are able to judge which information it is acceptable to disclose [5]. Information that is perceived to have a lower level of risk is regarded as having lower sensitivity and users are more willing to disclose this type of information compared to information that is perceived to have higher sensitivity [29]. Generally, information that can increase individuals' vulnerability is perceived to be more sensitive due to potential losses incurred. Among the potential losses due to disclosure of personal information are psychological, physical and material losses [24]. Also, information that relates to or can identify individuals is perceived to have higher sensitivity. Thus, sensitivity to information can be defined as the degree of privacy concerns towards certain information in a specific situation [34].

2.2 Disclosure in OSN

While many different kinds of OSN have emerged, the basis of an OSN is that it has visible user profiles with contacts linking to them [4]. The OSN's profile usually contains identifying information about the users (e.g. name, photo), their contacts (e.g. friends or connected users) and interests. This profile will form an impression for other users, who will eventually use this information to assess the individual when deciding on whether to form a relationship. Different types of profile information have been found to have different outcomes on the number of listed contacts. Information that shares common referents (e.g. home town, institution) increases connections among users [18].

To determine what information could be disclosed on Facebook profiles, Nosko et al. [26] conducted a content analysis on 400 randomly selected Facebook profiles. Out of 100 different types of information that were identified, four were related to employment information. First is information about the individual's *employer*. This information was disclosed publicly by 35.3% of the sample. Second is information about *current or previous job*. This information was disclosed by 32.5% of the sample. The third piece of information is the *working position* (disclosed by 30.5% of the sample) and the fourth is the *job description* (disclosed by 17.80% of the sample). Based on these results, most Facebook users prefer not to publicly share their employment information when they are given the option to do so. This could be because this kind of information can be used to locate and identify an individual and can further be used by others for malicious purposes. Compared to Facebook users, more than half of the blog authors surveyed in a study by Herring et al. [14] stated their employment information on their publicly viewable blogs. In their survey, the majority of the sampled blogs were created for personal purposes.

Employment information is categorized as sensitive information, along with (for example) profile pictures, photo albums, viewable friends, email address, relationship status and medical and criminal records [2, 26]. It is rated as medium in terms of sensitivity in e-commerce settings [21]. In contrast, except for email address (i.e. 43.3%), other information that was categorized as sensitive information (similar category to employment information) was revealed by more than 70% of the sample [26]. Thus, employment information has a unique position among Internet users, being the least shared sensitive information on users' OSN profiles.

Facebook users were found to adopt high self-censorship practices when disclosing publicly information on their profiles. Nosko et al. [26], in their investigation of information disclosure in OSN profiles, selected Facebook as their chosen OSN, as is one of the most popular OSNs available. From their sample of 400 Canadian users, they discovered that individuals only choose to publicly reveal 25% of the total available information fields on their profiles. This indicates that users are using a high degree of control in limiting access to their personal information. Similarly, when users were given the option to complete their profiles, slightly less than half of the profile fields were left unfilled [18].

2.3 Disclosure by Organization

Public organizational websites were created to serve as an official contact point between the public and the government. In the e-government concept, publishing organization information assists in increasing transparency and accountability and reducing corruption [25]. As well as public organizational websites [35], educational based organization websites also publish their employees' employment information [12]. Employees' information is published along with their employment details in order to offer better information and higher quality of services to the public when dealing with these organizations.

While individuals themselves disclose employment information on OSN, they are not directly responsible for disclosure on organizational websites. Therefore, individuals have less control over the disclosure of employment information on organizational websites compared to OSNs, blogs or personal websites. Questions are also raised as to what kinds of information organizations should be allowed to publish. Can any information (even if relevant to the organization) be published online if it makes the employees uncomfortable?

This study will draw from information boundary theory to compare individuals' behaviour and perceptions towards employment information disclosure on two different platforms. It attempts to investigate how individuals feel and their experience when their employment information is published on the Internet by focusing on disclosure through organizational websites and disclosure on an OSN, i.e. Facebook.

3. INFORMATION BOUNDARY THEORY

Information boundary theory (IBT) suggests that individuals' reactions to the collection of their information should follow rules for "boundary opening" and "boundary closure" [31]. IBT theory is synthesized from communication boundary theory, justice theory and a general-expectancy-valence framework for privacy protection. IBT is also known as communication privacy management theory (CPM: [27]) due to the similarity in the boundary metaphors to explain the privacy management process [19]. Initially CPM was known as Communication Boundary Management Theory, with the intention to examine factors for individuals' privacy management with respect to specific relationships [28]. Eleven years later, the term was adjusted from Communication Boundary Management to Communication Privacy Management to emphasize the key privacy issues undergirding the theory. Although early research on CPM was conducted within interpersonal relationships, the theory has recently expanded to explain disclosure within online information privacy research [15, 23]. In his review of established theories in online information privacy research, Li [19] identifies IBT as one of the established theories in privacy research that have been empirically tested at individual level. Further, IBT has also been applied to employee privacy research [32], which is the focus of this study.

Information boundary theory (IBT) suggests that each individual develops a boundary coordination process around an informational space through which the person manages the decisions to reveal or withhold information via a given medium. These information boundaries are based on certain conditions or rules that individuals develop to assist them in making decisions [31]. Information is released when the boundary is open and withheld when the boundary is closed. Boundary rules are applicable in online interaction and observational forms of media. These boundaries assist individuals in controlling access to information and judging expectations for mutual information ownership [27]. In this view, the privacy regulations of individuals as well as others who are granted access to the individual's information (the flow of information) are considered. The negotiation of boundaries (i.e. strict or loose) is dynamic depending on the situational context, e.g. level of risk related to information privacy. For example, the higher the perceived risk, the stricter the boundary. In the online environment, the risk of disclosing (personal) information could have undesirable consequences such as embarrassment, vulnerability and financial loss. The theory proposes that individuals develop three general principles in the management process when deciding on boundary regulations in order to protect personal privacy. First, 'boundary rule formation' refers to the individual's ability to control how much information is revealed or withheld across boundaries. Second, 'boundary coordination' stipulates that individuals enact rules based on expectations of information usage and access to the information outside the boundary. Since boundary management processes involve both individual (i.e. personal) boundaries and collective boundaries (private information co-owned with other parties), boundary coordination between co-owners of information is negotiated to determine privacy access and privacy protection. Third, 'boundary turbulence' occurs when the boundary coordination process is unsuccessful. When this happens, boundary coordination fails to work and individuals or co-owners will seek remedial action to restore boundary management to an acceptable level. These boundary rules are influenced by, for example, the nature of the relationship, information usage, and the benefits of disclosure [27].

According to Petronio [27], there are five principles underpinning the rule management system: these are ownership, control, privacy boundaries, co-ownership and privacy turbulence. People believe that they own their personal information, and therefore that they are entitled to control the flow of this information and make decisions about it based on privacy boundary rules. The information that they share with co-owners is deemed to establish acceptable privacy rules, and when privacy rules are disrupted or violated, this may result in privacy turbulence.

This study applies IBT to public employees to understand personal information disclosure in two different platforms, i.e. OSN and official websites. This disclosure is focusing specifically on OSN profiles and organizations' official websites. It will focus on how public employees enact boundary rules when a particular attribute about them is disclosed in two different contexts. It is argued that Internet users' boundary decisions are based on situational context [16]. In addition, it is also suggested that the status of the relationship between the sender and the receiver (individual or institutional) may have an influence towards articulating privacy boundaries [32]. This paper applies IBT to personal information (i.e. employment information) disclosure and adds our understanding of contextual online disclosure.

4. METHODOLOGY

4.1 Semi-Structured Interviews

To explore participants' views when their employment information is published on the Internet, this study employs a qualitative strategy, i.e. face-to-face semi-structured interviews. A qualitative strategy was chosen because this approach can provide an understanding of the context or settings where the participants in a study address an issue and can provide richer insights into participants' voices, stories, viewpoints and feelings [7]. The interviews were conducted in Malay, which is the official language in Malaysia and is used in official government communications. A semi-structured interview technique was used because of its flexibility in posing additional questions to gain opportunities to identify new themes or ideas to clarify or illuminate the research problem. In addition, to gain information about participants' working experience and background, interview questions focused on the disclosure of their employment information on the Internet, specifically on their organization's website, and on their personal online social network profiles. In order to gain participants' views on disclosure of employment information, participants were asked twice about this in different

stages of the interview. During the early part of the interview, participants' perceptions regarding disclosure of employment information were explored in general. Questions were posed as to how they perceived employment information disclosure on official websites. Then questions were asked focusing on their OSN accounts and profiles. Their behaviour towards publishing their employment information was also investigated. Towards the end, participants were asked again about their perceptions about disclosure. Participants were also asked about their feeling and experiences on both occasions.

4.2 Participants

A purposive sampling technique was adopted in this study to identify the participants. Participants were recruited on the basis of their working experience in the Malaysian public sector and their participation in online social networks. It was important that participants' organizations had official websites in order to understand how they viewed disclosure on these websites. Similarly, participants' participation in OSN will lead to an understanding of disclosure in a social context to enable further analysis of both contexts.

Potential participants were contacted via phone, email and OSN, i.e. Facebook. The purpose of the study was explained to the participants, albeit in general terms, during the first contact. Ethical approval was obtained from the university ethics committee and participants' consent was sought prior to the interviews. The consent document reiterated details of the study and protection of the participants' data. Five participants from Malaysia with between six and ten years' working experience in the public sector were interviewed. All participants were aged between thirty-one and forty years and were from the professional and management category, with positions as researchers, science officers and administrative officers. Four participants were female while one was male, four participants were married and all participants had received tertiary education. Participants were culturally and nationally homogeneous, allowing further understanding of the cultural and national context.

4.3 Data analysis

All interviews were audio recorded and transcribed by the first author. Transcription was carried out using NVivo version 10 software and data were further coded using the same software. Transcription data were translated into English by an English language lecturer currently pursuing a PhD in the UK. Data were coded based on themes derived from a literature review (*a priori* coding – codes were developed before examining the current data) and open coding, which is based on data that emerged from the interviews (inductive coding). Coding was conducted by the first author. The coding process involved identifying categories, patterns and themes on participant's experiences and looking at how they linked each other.

The interviews were analyzed using thematic analysis. This approach was chosen because it allows the researcher to identify factors or variables that emerge from the participants [8] and data can be displayed and classified according to similarities and differences. Thematic analysis can be used for reducing and managing large volumes of data, getting immersed in the data, summarizing the data and focusing on interpretation. Coded data were grouped into similar categories and were clustered into themes.

5. RESULTS

5.1 Disclosure on Official Websites

Overall, participants in this sample were aware of their employment information disclosure on their official organizations' websites. Participants were able to confirm instantly when asked about this. They listed working position, working grade, job role, work promotion, work transfer and employer's information (i.e. organization, address). Participants admitted that they themselves had done some cross-checking of their published information. They did this either by browsing the organization's website or by using a search engine. This was to ensure the accuracy of their information in order that they could inform the relevant unit if any information was inaccurate. In addition, all participants reported that they had Facebook accounts and were active OSN users.

Participants generally felt that publication of their employment information on their organizations' public websites might increase their reputation. Participants felt proud and excited when they noticed that their employment information was published on their organization's official website. One possible reason was that working with the government is considered as an honour for the individual and as a way to serve the country. Another reason might be because a career in government is generally respected, since it is difficult to secure such a position and the competition is very tough. One participant viewed it as an honour to serve the government, while another participant saw it as a responsibility:

> *"feel sort of proud (because) you are being someone in your country. …It is like at least I am doing something (for the country)."* (P1)

However, the feeling of pride and excitement reduces over time. Two participants mentioned that their feelings about the disclosure were not the same as when they started work. They characterized it as normal and not as something to be proud of.

Most participants perceived that disclosure is important for improving service delivery. It allows the public to contact them regarding their jobs. It serves as a means of identification for the public in finding the relevant position holders to seek feedback or assistance. For example, one participant from the finance unit explained:

> *"…so whatever they wanted to ask, they can come direct to us…since salary is a sensitive (issue), haa everyone want to email…as you know, who doesn't want to know their salary?"* (P2)

Another participant from a research institution stressed that employment information, especially one's position and expertise, is important for career development. This information will act as a marketing tool to enable employees to attract research collaboration from outside the participant's organization.

5.1.1 Privacy concerns

During this first round of questioning, while most participants expressed the importance of disclosure, one participant had a mixed reaction towards it. This participant expressed concern because of the ability to be contacted. While this might be of benefit to certain members of the public, the participant felt uncomfortable and explained her experience when one of her friends (from a different office) mentioned finding her on her organization's website. She was surprised because she did not want her (normal) friend to know more about her.

During the second round of questioning, the same question was asked again after exploring participants' behaviour and practice on their OSN accounts. The OSN questions brought up questions regarding privacy issues and how participants dealt with them. Therefore, by asking the same question at two different stages during the interview, a clearer understanding of the issue can be obtained.

Most participants seemed to have some concerns about employment disclosure compared to the earlier answers. An employee from the finance unit had previously described such disclosure as important; however, she showed some concerns during this second stage. She highlighted the experience of her colleague, whose ex-husband was able to find her based on the information on her organization's website. Although the wife had requested a transfer in order to distance herself from her ex-husband, she found this difficult to achieve, as her employment information was publicly available on her organization's website. Therefore, in this case, if anyone is trying to hide or run away from someone, the employment information that is published on the official website may provide an indication of their location. Moreover, the information on the official website is publicly viewable by anyone. Based on this information, employment information on organizations' websites could be misused as one way to locate the whereabouts of an individual. This traceability function was also highlighted by another participant in expressing the real-life danger of a government employee.

> *"As an example, recently a deputy director general was assassinated: we don't know whether the criminals might get that information that he is a Deputy Director from (official) website…we don't know either."* (P3)

Likewise, another participant clearly mentioned that working grade should not be disclosed to the public. Working grade is assigned to employees and determines the pay level for the job. It also describes the working category of the employee and provides general information on his or her organizational level. In view of this, the participant added that working position should also be concealed from public view.

> *"Although work position is general, others, especially those in the government circle, will know that an assistant secretary is at least grade 41"* (P5)

The participant above went on to explain the risk of revealing work position by specifically referring to high ranking government employees.

> *"…normally for those in the higher ranking position… might give the impression of a wealthy person. Could be a target for criminals"* (P5)

The responses underlined how important employment information is to employees. Revealing employment information may pose threats to their offline world. Reference can be made to the position and/or working grade, which will provide some indication of the social status of an employee. This could attract interested parties to target potential employees for financial crime.

5.2 Disclosure on OSN i.e. Facebook

All participants were active Facebook users. While most of them had more than one OSN account, all participants used Facebook more compared to other OSNs. Interestingly, none of them subscribed to LinkedIn. Although Facebook generally requires users to use their real identity, all participants reported using fabricated identities rather than their real names. Despite this, most of the participants chose a name related to their real names. Only one participant used a totally different name. One employees who had decided to used non-real name explained,

> *"If (anyone) wanted to search for others simply by typing a name, haa – you are found. Me, no…I don't want to be found."* (P3)

Another participant, who also used a pseudonym, believed that this practice would protect him from being interrupted or disturbed by people he did not know. He wanted to make sure that only those who really knew him would make contact with him and become his friend.

> *"I don't want any interruptions from unknown friends, those strangers"* (P5)

In addition to the practice of using pseudonyms as a way to seek privacy by hiding their identity, all participants had configured their Facebook profile privacy settings to private. Participants were aware that others could potentially view their profile content and were not comfortable with it. They took active steps to protect their profiles by limiting what others could see and disguising their profile names. It was evident that participants in this sample had some degree of privacy awareness and privacy concern.

While employment information on Facebook profiles is optional, Facebook encourages its users to complete all their profile information, including employment details. Out of the five participants in this study, three did not disclose any employment information on their profiles, while one participant only disclosed the (federal) department where she worked. Data from participants suggests that privacy concern was their main reason for not disclosing employment details.

> *"…because I don't want others, err those who don't know me, to know more about me"* (P5)

Additionally, participants viewed OSN as a tool for social purposes. Their main purpose of using Facebook was to keep in touch with friends, especially distant friends, maintaining relationships and getting updates on their friends' or Facebook groups' activities. A participant who is a researcher explains,

> *"For me, Facebook is more on social activities that are outside work. Because even for me, if anyone who is working in the same place asked to add me on Facebook, I normally won't approve their request, because for me, I think there are differences."* (P4)

Regardless of their privacy settings on their Facebook accounts, when given the option not to disclose their employment information, participants chose not to. These findings indicate that they preferred to keep employment information, even if protected (via privacy settings), to themselves rather than disclosing it. For them, this is an important piece of information and not to be shared even with their circle of friends.

6. DISCUSSION

The findings of this study describe public employees' behaviour and perceptions towards their employment information on two different platforms. Most striking was the way in which contextual factors influence boundary formation even within the same online environment.

Generally, participants had high privacy concerns with regard to their OSN accounts. All participants had configured their profile

settings to private to limit what others could see. When discussing employment information, participants perceived that this information is always linked with their personal attributes on both platforms. Employment information *per se* did not have any meaning for them if it could not be linked to their identity.

Employment information on organization websites was generally accepted by all participants. Participants generally had no objection when their employment information was published on organizational website. It provides some advantages to them and also to their organization. However, when issues of privacy were brought up indirectly, participants were found to construct boundaries in order to protect their employment information. Participants strictly limited the usage of their information to official purposes only. They were concerned that their information might be used for unintended purposes, with negative consequences for them. In short, the motivation to protect their information stemmed from closing boundaries to avoid negative outcomes [31].

On another note, participants showed different behaviour towards this information on their OSN profiles. Participants did not disclose their employment information when they had the option to do so. And when they did disclose this information, it was limited to their friends only. This indicates that employment information has high sensitivity among the public employees. The more sensitive the information, the stronger the decision not to disclose it [6]. The sensitivity of this type of information depended on the context. When employees perceived that official websites published official information for official purposes, they were willing to allow it to be published online in order to serve the public. However, participants did not disclose this information on their Facebook profiles. The situational difference between the professional and the social context led to their decision to create this boundary. Since participants perceived that OSN was for social purposes only, they did not disclose their employment information because they did not want to mix their social with their professional lives. Participant felt that this information was more sensitive when they had control over it but less sensitive when they did not. On Facebook, users have a high level of control and choice in deciding on disclosure of their personal information, compared to low control on their organizations' websites.

By making the decision that OSNs are for social purposes and official websites are for official purposes, participant have constructed a cognitive boundary between these two domains. Most participants made reference to their roles in each domain when deciding to create boundaries. In addition, this cognitive boundary is based on the activities performed by the participants in each domain. Despite having the same attributes in the same online environment, it creates a different level of sensitivity based on the context. This finding support previous research in determining that the individual's willingness to disclose personal information is based on contextual factors [16].

However, the degree of sensitivity was different when employees looked at the issue from the organizational website's perspective. Although participants mentioned their responsibility as government employees in order to justify their employment information being published, they also showed concerns about this publication (on the website). While they kept this information private on their ONS profiles, participants were not able to do so on the official website. One explanation for employees' perception in compliance with privacy boundaries is that this could be seen as loss of control over personal and collective boundaries. Participants considered that they have no control towards their employment information on their organizational websites, as it is up to the organization to decide what to publish and how. It can be observed that there is a tension between employees' interests in privacy and organizational interests in meeting objectives.

While control helps to preserve ownership, loss of control means loss of ownership of information. Individuals assume that they own their personal information. According to Petronio [27], individuals "have the right to own private information, either personally or collectively" (p. 6). In boundary setting, individuals decide what to divulge and what to withhold on the basis of ownership of information. When public employees refer to employment information on official websites, they believe that this information is owned by their organization. Hence, it is not owned by them, and they thus behave differently compared to when they believe they have ownership of the information. This was clearly shown when most of the participants chose not to disclose their employment information on their Facebook profiles.

7. CONCLUSION

By drawing on information boundary theory, this study helps to understand the decisions that public employees make about the disclosure of employment information in two different contexts/situations. Results suggest that participants create boundaries by differentiating between official/formal and social situations. This results in participants perceiving that control and ownership of employment information are different in those environments. In addition the degree of sensitivity of information, albeit the same information, depends on how the participants perceive the context. Nevertheless, this study is in its early stages and further investigation is required.

This study is not without limitations. First, only a small sample was recruited and the demographic properties of the participants were very similar in terms of working experience, age and qualifications. This could influence the findings due to an idiosyncratic phenomenon. Secondly, the study has limited the individuals' experiences to only one online social network, i.e. Facebook. Nevertheless, the popularity of Facebook has helped with the availability of data from our participants, as all of them have Facebook accounts. Research across different types of organization might have resulted in more conclusive data.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge Muhammad Yasir Yahya for his contribution to the paper.

9. REFERENCES

[1] Acquisti, A. & Grossklags, J., 2004. Privacy attitudes and privacy behavior: Losses, gains and hyperbolic discounting. In J. Camp & R. Lewis, eds. *The Economics of Information Security*. New York: Springer, pp. 1–15.

[2] Aimeur, E. & Lafond, M., 2013. The Scourge of Internet Personal Data Collection. In *2013 International Conference on Availability, Reliability and Security*. IEEE, pp. 821–828.

[3] Berendt, B., Günther, O. & Spiekermann, S., 2005. Privacy in e-commerce: Stated Preferences vs. Actual Behavior. *Communications of the ACM*, 48(4), pp.101–106.

[4] boyd, D.M. & Elison, N.B., 2007. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), pp.210–230

[5] Bryce, J., 2008. Bridging the digital divide: Executive summary, London.

[6] Castañeda, J.A. & Montoro, F.J., 2007. The effect of Internet general privacy concern on customer behavior. *Electronic Commerce Research*, 7(2), pp.117–141.

[7] Creswell, J.W., 2013a. Qualitative Inquiry and Research Design: Choosing Among Five Approaches Third Edit., Sage Publications Inc.

[8] Creswell, J.W., 2013b. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* 4th ed., Thousand Oaks US: Sage Publications Inc.

[9] Culnan, M.J., 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes Toward Secondary Information Use. *MIS Quarterly*, 17(3), pp.341–363.

[10] van Dijck, J., 2013. "You have one identity": performing the self on Facebook and LinkedIn. *Media, Culture & Society*, 35, pp.199–215.

[11] Fox, S. et al., 2000. Trust and privacy online: Why Americans want to rewrite the rules. *The Pew Internet & American Life Project.*, pp.1–29.

[12] Gallego-Álvarez, I., Rodríguez-Domínguez, L. & García-Sánchez, I.-M., 2011. Information disclosed online by Spanish universities: content and explanatory factors. *Online Information Review*, 35(3), pp.360–385.

[13] Gross, R. & Acquisti, A., 2005. Information Revelation and Privacy in Online Social Networks (The Facebook case). In *In Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*. WPES '05. Alexandria, Virginia, USA: ACM Press, pp. 71–80.

[14] Herring, S.C. et al., 2004. Bridging the gap: a genre analysis of Weblogs. In *37th Annual Hawaii International Conference on System Sciences*. Big Island, HI, USA: IEEE, pp. 1–11.

[15] Ji, P. & Lieber, P.., 2010. Am I safe? Exploring relationships between primary territories and online privacy. *Journal of Internet Commerce*, 9(1), pp.3–22.

[16] John, L.K., Acquisti, A. & Loewenstein, G., 2011. Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information. *The Journal of Consumer Research*, 37(5), pp.858–873.

[17] Labitzke, S., Taranu, I. & Hartenstein, H., 2011. What your friends tell others about you: Low cost linkability of social network profiles. In *Proceedings of 5th International ACM Workshop on Social Network Mining and Analysis*. San Diego, CA, USA.

[18] Lampe, C., Ellison, N. & Steinfield, C., 2007. A Familiar Face(book): Profile Elements as Signals in an Online Social Network. In *CHI'07 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California, USA: ACM, pp. 435–444.

[19] Li, Y., 2012. Theories in online information privacy research: A critical review and an integrated framework. *Decision Support Systems*, 54(1), pp.471–481.

[20] LinkedIn, 2015. About LinkedIn. Available at: https://press.linkedin.com/about-linkedin [Accessed June 15, 2015].

[21] Lwin, M., Wirtz, J. & Williams, J.D., 2007. Consumer online privacy concerns and responses: a power–responsibility equilibrium perspective. *Journal of the Academy of Marketing Science*, 35(4), pp.572–585.

[22] Madden, M. & Rainie, L., 2015. *American' Attitudes About Privacy, Security and Surveillance*, Available at: http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/.

[23] Metzger, M.J., 2007. Communication privacy management in electronic commerce. *Journal of Computer-Mediated Communication*, 12(2), pp.1–27.

[24] Mothersbaugh, D.L. et al., 2012. Disclosure Antecedents in an Online Service Context: The Role of Sensitivity of Information. *Journal of Service Research*, 15(1), pp.76–98.

[25] Ndou, V., 2004. E-government for Developing Countries: Opportunities and Challenges. *The Electronic Journal on Information System in Developing Countries*, 18(1), pp.1–24.

[26] Nosko, A., Wood, E. & Molema, S., 2010. All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, 26(3), pp.406–418.

[27] Petronio, S., 2002. *Boundaries of Privacy: Dialectics of Disclosure*, State University of New York Press.

[28] Petronio, S.S., 1991. Communication boundary management: a theoretical model of managing disclosure of private information between marital couples. *Communication Theory*, pp.311–335.

[29] Phelps, J., Nowak, G. & Ferrell, G., 2000. Privacy Concerns and Consumer Willingness to Provide Personal Information. *Journal of Public Policy & Marketing*, 19(1), pp.27–41.

[30] Smith, H.J. et al., 1996. Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS Quarterly*, 20(2), pp.167–196.

[31] Stanton, J.M., 2003. Information Technology and Privacy: A Boundary Management Perspective. In S. Clarke et al., eds. *Socio-Technical and Human Cognition Elements of Information Systems*. Information Science Publishing, pp. 79–103.

[32] Stanton, J.M. & Stam, K.R., 2003. Information Technology, Privacy, and Power within Organizations: a view from Boundary Theory and Social Exchange perspectives . *Surveillance and Society*, 1(2), pp.152–190.

[33] Stutzman, F., Gross, R. & Acquisti, A., 2013. Silent Listeners: The Evolution of Privacy and Disclosure on Facebook. *Journal of Privacy and Confidentiality*, 4(2), pp.7–41.

[34] Weible, R.J., 1993. *Privacy and data: an empirical study of the influence and types and data and situational context upon privacy perceptions.* Mississippi State University. Cited by Sheehan K., & Hoy, M., 2000. Dimensions of Privacy Concern among Online Consumers. Journal of Public Policy & Marketing, 19(1), pp.62–73.

[35] Zhao, J.J. & Zhao, S.Y., 2010. Opportunities and threats: A security assessment of state e-government websites. *Government Information Quarterly*, 27(1), pp.49–56

Boundary Enforcement and Social Disruption through Computer-Mediated Communication

Alexis Elder
Philosophy, Southern Connecticut
State University
501 Crescent Street
New Haven, CT 06515 USA
+12033926794
alexis.elder@gmail.com

ABSTRACT

In this paper, I investigate the impact boundary-promoting communication technology – such as texts, comments, microblogging, and instant messaging – have on friendships, and arrive at the surprising conclusion that these technologies are, despite appearances, good for personal relationships, and thereby good for us.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues – *ethics, privacy, use/abuse of power*.

General Terms

Design, Human Factors.

Keywords

Friendship; social media; computer-mediated communication; autonomy; virtue ethics.

1. INTRODUCTION

In her book *Alone Together*, Sherry Turkle reports that sixteen-year-old Audrey, like many of her cohorts, much prefers texting and other asynchronous communication methods to real-time communication methods, such as the telephone.

> "The phone, it's awkward. I don't see the point. Too much just a recap and sharing feelings. With a text… I can answer on my own time. I can respond. I can ignore it. So it really works with my mood. I'm not bound to anything, no commitment… I have control over the conversation and also more control over what I say."

Audrey proceeds to explain what she prefers about her favored communication channels, coining a new word in order to better articulate her reasons. Over the telephone, she says, "there is a lot less *boundness* to the person" than when texting. By this, she seems to mean that her ability to control what she says, and when, is greatly diminished by phone as opposed to text. On a phone call, a person can demand of her things she is not comfortable giving. Callers seem to her to be imposing on her time, and even her psychological space. Turkle's diagnosis is that, for Audrey, "A call has insufficient boundaries." [15]

In this paper, I investigate the impact boundary-promoting communication technology, such as texts, comments, microblogging, and instant messaging have on friendships, and arrive at the surprising conclusion that these technologies are, despite appearances, good for personal relationships, and thereby good for us.

My strategy is as follows. First, I investigate Friendship and the Good Life. I give reasons to think friendship is important for good human lives, and sketch features that make some friendships better than others. Then, I review Problems with Computer-mediated communication (CMC). Following that, I connect boundary-promoting properties of CMC and Autonomy. In that section I evaluate Marilyn Friedman's analysis of autonomy as a factor for social disruption, in order to argue that although autonomy is disruptive, it is potentially good-making for the best and most valuable relationships. I apply this analysis of Relationship Disruption and Relationship Quality to the boundary-promoting aspects of CMC by examining the role of Healthy Boundaries in Healthy Friendships. I conclude by noting that some of the potential shortcomings of CMCs, which remain despite the good-making features I have identified, are partly mitigated by technological solutions, most prominently in the increasingly popular app Snapchat. This may help explain the appeal of Snapchat, and suggests that we have reason to be sanguine about the long-term impacts of CMC. People seem to intrinsically value personal relationships, and to be eager to adopt technologies that promote good relationships.

2. FRIENDSHIP AND THE GOOD LIFE

In the opening passage of Book XIII of the *Nicomachean Ethics*, Aristotle claims that "no one would choose to live without friends even if he had all the other goods" [1]. Not everyone concurs with this sweeping generalization, but even for those who do not find his formulation plausible, friendship is generally taken to be something that makes a good life go better. We are social animals, and friendship is an important ideal of sociality.

The concept of friendship is difficult to pin down precisely, but there are several features worth noting. First, friendship is necessarily reciprocal. It seems to be a category mistake to talk about unrequited friendship, even though other positive dispositions toward people, such as love and affection, do not suffer from such problems. Second, friendship overlaps with other kinds of relationships. One can consider a sibling to be a friend.

{Publication}, Month 1–2, 2015, City, State, Country.

Likewise, one can be friends with a coworker, or a neighbor. Furthermore, to say of some relationship that it is (also) a friendship contributes an additional positive evaluation to the relationship – to be friends with one's coworkers is to say something good about the relationship, and to say that one is not friends with a sibling is to say something bad about it.

We sometimes use the word "friend" loosely, to refer to everyone with whom we are occasionally friendly. We also use the term more strictly, as when we speak of someone's being revealed to be a "true friend". In what follows, I will draw primarily upon Aristotle's theory of friendship. Aristotelian ethics offers some distinct advantages in the field of computer ethics, such as its focus on the ways that even seemingly small changes to how a person lives her life can, when exercised repeatedly, impact character [16]. In addition, Aristotle offers a robust theory of friendship that offers rich conceptual resources to draw on.

Aristotle distinguished between three major kinds of friends: utility friends, pleasure friends, and virtue or character friends [6]. The first two kinds are, as their names suggest, primarily instrumental relationships, rooted in the logically independent goods such people can provide each other, such as pleasure and utility: that is, goods that are not conceptually bound up with a particular person but could be provided by other people. My friend may be useful to me because he owns a big truck and can help me move bulky items, but anyone else with a similar truck would be similarly useful. The last form of friendship is considered by Aristotle to be the best, and is characterized by people valuing each other for their good character, their intrinsic qualities, as ends in themselves rather than means to other ends. This leads Aristotelian theorists such as Neera Badhwar to characterize the best friendships as "end friendships." Badhwar argues that end friendship is marked by "necessary irreplaceability" of the friend. "In an end friendship," she explains, "one loves the friend as an essential part of one's system of ends and not solely, or even primarily, as a means to an independent end… In such love, one loves the friend for the person that she is". End friends, she argues, are constitutive goods, not necessarily maximizing goods. One could believe, of a friend, "I might well have been happier on the whole without this friendship, whose presence is now a unique and irreplaceable constituent of my good" [2]. We thus have the resources to make senses of both loose and strict senses of friendship, as well as a partial theory of what makes some friendships better than others. In what follows, I focus on CMCs' impact on end friendship, as the best and fullest form of friendship, and so the closest to our ideals about what constitutes the highest-quality personal relationships.

End friendship can seem to be promoted by many forms of computer-mediated communication (hereafter CMC), from instant messaging, to email, to texting, to social media ranging from Facebook to Snapchat. It enables friends to keep in touch with each other even when they are physically distant, and while some (such as Skype and FaceTime) are synchronous, many others are asynchronous. They allow people to converse even when their schedules do not align. Imagine two friends, Alice and Betty. Alice works at a bank during normal business hours, while Betty's schedule as a nurse varies wildly from day to day, with occasional overnight shifts. Alice and Betty can send each other messages using CMCs whenever they have a spare moment, and be assured that their messages will be picked up by the recipient when the individual has time and energy to respond. Unlike older technologies, such as telephones, many CMCs thus permit people with widely varying schedules to remain in contact with each other, regardless of their physical proximity.

3. POTENTIAL PROBLEMS WITH CMC

Despite these apparent advantages, some aspects of CMCs seem to be problematic for friendship. I will focus on two major ones here. Both involve what we might think of as moral hazards – ways in which CMCs tempt us to engage in behavior that seems convenient at the time (and possibly even conducive to friendship), but that may have deleterious consequences for both individual relationships, and one's ability to successfully enjoy rich, rewarding friendships in the long-term.

The first concern involves the way that CMCs reward impatience, and thus seems to run the risk of making us worse friends. The worry is well articulated by Shannon Vallor [16, 17]. Vallor calls patience a "communicative virtue", a character trait that contributes to good, rewarding communication. Of course, many things do not cultivate the virtue of patience but that does not make them bad. Specifically, CMCs tend to reward impatience, thereby impeding the cultivation of patience. She argues that virtues are habits that are difficult, at first, to instill. With many CMCs, it is easy to exit a boring, tedious, or uncomfortable conversation – easier than it would be in a face-to-face exchange. One can click away from a chat window, swipe to another friend's page, or simply close a browser window or put away a smartphone without subjecting oneself to the kind of social pressure exerted when one's interlocutor is in the room. In fact, given the asynchronous nature of CMCs and the assumption that people will drift in and out of conversations as time permits, exiting for other reasons may be all but invisible.

The immediate appeal of CMC is that it lets us avoid difficult, uncomfortable situations, by clicking out of a chat window or navigating away from a blog post, or, as Audrey describes, by answering a text in one's own time… or not at all. Such actions may, however, harm friendships, even the best of which can require patience on occasion. And even when a particular friendship is undamaged by such actions, they give reason to be wary [17]. They cultivate deep-seated habits of avoiding difficult conversations, which in turn impedes the development of character and so our ability to live the good life with friends.

Computer-mediated communication's tendency to undercut patience may also interfere with our ability to cultivate empathy, another communicative virtue [16, 17]. Empathy requires willingness to sit through difficult experiences – in other words, patience, making CMCs potentially doubly damaging because they can weaken both patience and empathy. Patience is, furthermore, important for establishing that one values the whole person in a relationship, and not merely the good things they provide. Vallor notes that patience, once established, facilitates interpersonal trust by "communicating to others that your interest in them does not end with their ability to keep you constantly pleased or fascinated." [17] Impatience with others contributes to the impression that the goods one gets from one's friends are primarily instrumental, such that one will opt out when immediate tangible goods such as pleasure or utility are lacking. Patience and empathy, by contrast, show that one is concerned with - and values – the whole person, for themselves and not merely for the external goods they provide.

This worry about instrumentality is also expressed by Turkle. "Networked, we are together," she argues, "but so lessened are our expectations of each other that we can feel utterly alone. And there is the risk that we come to see others as objects to be

accessed—and only for the parts we find useful, comforting, or amusing" [15]. This risk stems, she argues, from the ease with which CMCs allow us to cycle through our social network, in search of gratifying contact, while avoiding those who do not give us what we want, when we want it. She offers as evidence a conversation between herself and fifteen-year-old Ricki:

> "I have a lot of people on my contact list. If one friend doesn't 'get it,' I call another." This marks a turn to a hyper-other-directedness. This young woman's contact or buddy list has become something like a list of "spare parts" for her fragile adolescent self. When she uses the expression "get it," I think she means "pick up the phone." I check with her if I have gotten this right. She says, "'Get it,' yeah, 'pick up,' but also 'get it,' 'get me.'" Ricki counts on her friends to finish her thoughts…. [15]

This impatience with those who do not immediately "get it" can lead, Turkle fears, to the illusion that friends are good merely for scratching one's social itches.

> In a life of texting and messaging, those on that contact list can be made to appear almost on demand. You can take what you need and move on. And, if not gratified, you can try someone else. [15]

By making it easier to "use" friends to satisfy needs and then put them aside with little or no cost when they ask for something in return, CMC may have a tendency to reinforce in us habits of treating friends as replaceable sources of repeatable goods, rather than irreplaceable constituents of our good, as the best form of friendship seems to require. Features that make it easy to treat friends as interchangeable, always available (in some sense) and easily dismissed when what they offer is not quite what we're looking for, seem as though they are detrimental to friendship.

4. CMC AND AUTONOMY

Despite these concerns, I think that CMCs are, on balance, beneficial to developing and maintaining high-quality friendship. In order to see why, I first discuss the ultimate source of worries about both patience and instrumentality. They are traceable to a specific kind of moral hazard: the ease with which a person can choose to both engage in and disengage from exchanges with others. As Vallor puts it, many CMCs implemented in popular social media programs offer "low entry and exit barriers, when compared with social environments (online and offline) with … higher entry/exit barriers" [16]. CMCs thus facilitate the enforcement of boundaries between individuals. It is one's ability to enter or exit at will that permits one to establish or maintain boundaries, to, as Audrey puts it, "have control over the conversation and also more control over what I say"[15].

Her appeal to control suggests a connection between such costs and personal autonomy. Lower barriers to and lower (external) costs of boundary establishment make it easier for individuals to choose whether to engage. People are not, as Audrey puts it, "bound to anything, no commitment…". Her decisions about whether, when, and with whom to interact are made at her own discretion, on the basis of her own reasons and not subject to oversight by others [15].

Justine Johnstone has argued that computer ethics can benefit from considering technology as empowerment [10]. With CMCs, lowering associated costs of engagement empowers individuals to enter or exit conversations – and relationships – at will. CMC users are, in this respect, capable of exercising more autonomy than in those where, as Vallor puts it, "situational opportunities… exert some pressure upon us," specifically "the social strains and burdens of face to face conversation" [16]. To some, this may sound like a fairly light pressure and so not a significant impediment to autonomy. But it has been documented (e.g. in [7, 14]) that these social strains are not equitably distributed: women, members of racial and ethnic minorities, young and disenfranchised people may experience disproportionate social pressures and have correspondingly more difficulty engaging and disengaging from social interactions at will, at least in face to face interactions. These people would thus be positioned to find CMC empowering with respect to their ability to exercise autonomy in social settings.

The fact that CMC empowers by enhancing autonomy does not, by itself, do much to mitigate the concerns raised earlier. Rather, it makes transparent the mechanisms by which they are created. In "Autonomy, Social Disruption, and Women," Marilyn Friedman uses the painter Paul Gauguin as an example of a person who, in exercising autonomy, disrupts social bonds. Gauguin abandoned his wife and children in order to pursue his career as a painter. While he might be lauded for his commitment to artistry, he seems badly lacking on the social front. Friedman points out that he exemplifies the potential of autonomy to disrupt personal relationships – much as Vallor and Turkle worry that widespread CMC use will be disruptive to friendships [9].

Friedman's analysis of the mechanisms by which the exercise of autonomy disrupts relationships sheds light on the impact of CMC on friendship.

"Whenever someone questions or evaluates any tie or commitment that binds her to others," says Friedman,

> the possibility arises that she may find that bond unwarranted and begin to reject it. Rejecting values that tie someone to others may lead her to try to change the relationships in question or simply to detach herself from them. Someone might also reflect on the very nature of her relationships to particular others and come to believe that those ties are neglecting or smothering important dimensions of herself. To liberate those aspects of herself, she might have to distance herself from the problematic relationships. [9]

Furthermore, points out Friedman, exercise of autonomy can also disrupt by making her a less desirable companion to others:

> …someone's increasing autonomy might result in the breakup of a relationship not because she rejects it but rather because other parties to the relationship reject her. They might despise the changes in her behavior that they are witnessing. Some parents, for example, disown children who rebel too strongly against deeply held parental values. Peer groups often ostracize their members for disregarding important norms that prevail in their own subcultures. [9]

Autonomy thus threatens relationship by giving one power and opportunity to evaluate whether or not to preserve any given social tie, and adds an additional layer of risk by introducing the potential to change values and behavior in ways that make her an undesirable companion to her friends. This corresponds closely to the thought that CMC use can encourage one to become impatient and thereby a worse friend, and also that in facilitating the switch from one friend in a contact list to the next, one's social ties may be weakened or cut altogether. The autonomy enabled by CMC can thus be socially disruptive.

5. RELATIONSHIP DISRUPTION AND RELATIONSHIP QUALITY

I conclude that critics are right to be concerned that social media may disrupt and weaken relationships, both by making exit easier, and by allowing users to choose which messages to respond to and when. However, the ability to exit, as well as to enforce boundaries, can at the same time be a relationship enhancer in two ways

First, autonomy-promoting aspects of computer-mediated communication also have the potential to reinforce relationships because one can choose to engage with people. Friedman points out that autonomy "might lead [a person] to appreciate in a new light the worth of her relationships or the people to whom she is socially attached and to enrich her commitment to them. In such cases, autonomy would strengthen rather than weaken relational ties." [9] As the story of Ricki, who would go through her contact list until she found someone who "got" her, inadvertently illustrates, we often give up on some interactions by exchanging them for others. Friedman points out that people often turn away from some communities as part of a process of turning toward others. "Without any empirical backing," she says, "I nevertheless estimate that in most cases in which autonomous reflection does lead people to reject the commitments that bound them to particular others, they are at the same time taking up new commitments that link them through newly shared conviction to *different* particular others." [9]

Friedman's point can be taken further. Suppose one believes that the best relationships are those in which all participants choose (or would freely choose) to participate. Suppose furthermore that relationships from which a person would choose to exit, given the opportunity, tend to be suboptimal. It would then follow that even though autonomy threatens particular relationships, it does so by a sort of filtering process, in which the less-than-ideal relationships tend to be disrupted, while the best relationships tend to be reinforced. That would not guarantee that every relationship disrupted would be one that was not worth having to begin with, nor that every relationship mutually chosen would be a worthwhile one, but it would tend over time to select better ones over worse.

This is a more sanguine picture of friend-replacement than that painted by Turkle. It does not alleviate the concern that CMC use leads to an instrumental attitude toward friendship. But it does suggest the need for a more fine-grained analysis that avoids a false dichotomy between forced togetherness and isolating autonomy.

The original concern was that the ability to switch out connections would tend to reinforce individuals' tendency to treat friends as interchangeable sources of repeatable goods, like pleasure and utility, rather than as irreplaceable constitutive ends in themselves. Given Aristotle's distinction between the highest form of friendship, character friendship, and the lesser instrumental kinds of pleasure and utility friendship, CMCs then appeared to promote lesser forms of friendship over higher. However, it also looks as though at least some exercises of relational autonomy can reinforce social bonds by affirming particular individuals as worthy of choosing. Furthermore, CMC users can exercise their autonomy to avoid people who make poor friends, but with whom they might otherwise engage, owing to social pressures. When individuals reciprocally choose to interact, because each finds the interaction choice worthy, it is not obvious that we should conclude that they do so on the basis of instrumental benefit. At least some may be chosen on the basis of more friendly criteria. It is far from clear that the choices made via social media are any more likely to be made for instrumental rather than intrinsic, character-driven reasons. In fact, while there may be an initial temptation to choose on the basis of instant gratification, to the extent that this makes a person less choice worthy to others, a better grasp of long-term reciprocity and the value of listening to others may be reinforced over time.

At minimum, it appears that the autonomy CMCs promote is that of choosing which out of an array of friends to engage with. The basis, on which the choice is made, however, is underdetermined by the evidence. Those who, like Turkle, worry that this leads to a kind of instrumentality seem to assume that (at least many of) the choices will be made on the basis of the pleasure or usefulness provided by different friends. But not all choices need be made for instrumental reasons – some could be on the basis of character, for instance. And not every relationship is worth sticking with. If CMCs present a moral hazard in making it easier to opt out of exchanges when they are unrewarding, they may also reinforce positive exercise of autonomy in opting out of interactions with poor companions. Thus, we have at least one reason to think that CMCs are potential good-makers for friendship.

A second reason stems from the use of CMC to enforce boundaries. Even should one choose to disengage from a particular exchange and use CMCs to enforce boundaries in a particular relationship, as when someone declines to answer a text immediately, this does not necessarily pose a threat to that relationship. I appeal to a theory of friendships as composite social groups to explain how such groups can be strengthened by what we can think of as healthy boundaries between friends. It is thus a potential good-maker for relationships and for us. Furthermore, these benefits can plausibly outweigh the risks previously identified.

6. HEALTHY BOUNDARIES IN HEALTHY FRIENDSHIPS

Thus far, I have focused on low-cost exits afforded by CMCs primarily as a matter of deciding whether or not a person is worth engaging with, either in the short- or long-term. This makes it seem as though exercise of autonomy is justified when a person's self-interest is in conflict with the wellbeing of some relationship. And sometimes this is the case.

But I am interested in exploring a less obvious way that low barriers to exit can enhance relationships – not merely as a sorting mechanism for screening out certain types of bad friendships (those we cannot get away from, or feel stuck with because no better option is available), but for improving the very friendships in which one occasionally exercises an option to exit.

To make sense of this, it will be helpful to return to the question of what makes a friendship good, and to reflect in somewhat more detail on Aristotle's account of the best kind of friendship.

Aristotle said, repeatedly and somewhat puzzlingly, that the best friends are "other selves" [1]. One way of reading this is that Aristotle thinks friends should be very similar to each other – mirror images or twins. (See Cocking and Kennett [5] for one such interpretation.) One benefit of this approach is that it is consistent with the fact that friendship networks tend to be highly homogeneous [11]. However, as a description of an ideal – what makes someone a *good* friend – it is lacking. One important good of friendship is the *different* perspectives and experiences friends

contribute [18], and idealizing similarity as a key quality of the best friendship rules out complementarity, the ways friendships can be enriched by differences between friends.

This problem can be avoided by adopting a different interpretation of what it means to be another self. Rather than think of friends as "other selves" because they are similar, we can start by thinking of friendships as composite objects in our social ontology. Friendships, like other close-knit social groups, can be thought of as being composed of people who – like parts of a machine or organism – may differ from each other but work together as an interdependent whole.[1] On this interpretation, friends' interactions would constitute their composing a friendship, and differences between friends, like differences between different parts, could actually enhance the functioning of the whole, allowing for a high degree of complementarity as friends contribute different strengths to their shared activities. The "parts" account of friendship would make friends out to be other selves by being other parts of the same whole. Aristotle's ideal, then, would turn out to prescribe interdependence and interaction without requiring similarity.

Asynchronous and spatially discontinuous exchanges enabled by CMCs are attractive to users precisely because they support interaction, and potentially interdependence, without requiring similarity of time or place. Thus, it can make it easier for people with different lifestyles and activities to keep in touch with each other. In this respect, the qualities that Audrey identified at the start – that she can choose when and whether to respond, according to availability and mood – look less sinister and more like features that can actually enhance very different people's ability to get along.

One reason that Audrey's comments and desire for control over her communications may look unfriendly is that there is a tendency to think that friendship involves dissolving of boundaries and merging of interests. Nancy Sherman, for instance, says of friendship that "it is a relaxing of one's own sense of boundaries and control. It is acknowledging a sense of union or merger" [13]. If this is correct, then Audrey's complaint that face-to-face conversation decreases boundaries is a complaint about a necessity for friendship. But the model I have just sketched, where friends are parts of the same whole, does not imply that boundaries between friends are inimical to friendship – rather, the reverse can be true. Internal boundaries between parts, to build on the metaphor, can lend structural integrity to the whole by enhancing each part's function on its own terms, so long as the part can continue to interact and react as appropriate.

This is consistent with how we see people actually using social media and other CMCs. Many teenagers today are occupied by a variety of extracurricular activities; their busy schedules and non-overlapping but time-intensive enrichment make it difficult for them to socialize in person. Although teen users of social media prefer to meet with their friends in person, they value social media because, it permits friends to keep in touch between classes, meets, rehearsals, and jobs [3]. Although one can read this as a cost of such intensive scheduling, it is also consistent with the thought that technology benefits by permitting greater engagement with projects without sacrificing one's relationships.

While it is tempting to idealize friendship as a selfless willingness to put one's own interests aside in order to listen to a friend, a more nuanced approach is called for. In order for friendships to be sustained and enriched by complementary companions, their complementarity may require assistance. Boundary protection can be a healthy part of mutual respect for divergent interests which themselves contribute to the value of the friendship. Vallor's point about the perils of rewarding impatience are still important. But CMCs can be beneficial to character by promoting a different kind of patience. Asynchronous communication reinforces the idea that different people are up to different things that require different schedules, and one needn't be the sort of person to put everything on hold for day-to-day conversation to be a good friend. The person with the ability to respond when her mood and attention allow her to read carefully and thoughtfully construct her answers, rather than whipping off a reply as soon as a message comes in, might well use this to be a better friend.

CMCs' ability to enforce boundaries, then, turns out to be a potentially good-making feature of personal relationships in several ways. It improves some by facilitating personal enrichment of friends. It improves individuals by reinforcing respect for others' time and projects. While it is potentially disruptive to relationships, its very potential to disrupt tends to suggest problems with such relationships to begin with. All other things being equal, CMCs are a qualified social good.

7. SNAPCHAT

While I have defended the claim that social goods follow from the increased autonomy associated with CMCs, there remains one further worry. CMCs empower individuals, but they sometimes seem to do so at the cost of equitability in relationship. Because, historically, CMC technologies have made most actions besides the completed communication opaque, they provide a kind of immunity from social judgment that may present another moral hazard. This hazard can be illustrated with a classic thought experiment from the philosophical canon.

In Book II of Plato's *Republic*, Glaucon challenges Socrates to defend the value of morality with a story about the Ring of Gyges, which turns its wearers invisible. Supposedly, formerly-virtuous individuals who wore the ring would behave reprehensibly once they put on the ring and could act with impunity, safe from scrutiny by others, stealing, committing adultery, and trespassing at will [12]. In many ways, CMCs function as communicational Rings of Gyges. When you send a message, you often have no idea when or whether it is picked up, whether the recipient skimmed and dismissed it or read carefully and sympathetically. "I didn't see the text" has become the 21st-century equivalent of "the check is in the mail." CMCs may thereby also present the same sort of moral hazard as the Ring of Gyges.

Furthermore, once a message is sent off, it becomes, in James Moor's memorable term, "greased information" - it may be forwarded, archived, posted en masse to social media or independent websites, and otherwise passed from hand to hand at lightning speed, difficult if not impossible for the sender to control.

These features may seem to contribute to the hazards identified earlier - reinforcing impatience as we flick through text messages and status updates, leading us to treat communications as commodities to be separated from their context when it is convenient to do so. But the rising popularity of a new social media - Snapchat - suggests that there is market demand for technologies that help us resist such hazards, preserving what is good about CMC while reducing some of the moral hazards of earlier technologies.

[1] I develop this theory in more detail in my dissertation, *Metaphysics of Friendship* [8].

Snapchat allows users to send pictures, often overlaid with scribbled artwork or text messages, to each other. Once a picture is received, it lasts for only a limited time - usually 10 seconds - and then deletes itself. Users can, of course, bypass this by various means, but if Snapchat detects that someone is trying to save the image, it sends a message to this effect to the original sender, thereby removing some of the invisibility of prior CMCs. To view the photo, the recipient must keep their finger on the screen, and upon opening a "snap", its sender is automatically alerted.

It is common to suppose that the bulk of Snapchat's appeal lies in its self-deleting photos, which facilitates sexting while mitigating some of the privacy concerns historically associated with this practice. But its teen user base seems to have embraced the app for much more broad usage [4]. Reflecting on the ways that Snapchat makes features of communication transparent that prior CMCs left opaque, and the nudges these provide, it is not implausible to suppose that part of the appeal is the way it mitigates some of the moral hazards associated with prior CMCs. As danah boyd puts it, "When someone sends you an image/video via Snapchat, they choose how long you get to view the image/video. The underlying message is simple: You've got 7 seconds. PAY ATTENTION"[4]. This message is enforced by the fact that viewing the message requires continued active engagement by the recipient. The message recipient's behavior is far less opaque than with previous technologies, removing some of the Ring-of-Gyges effect. This provides a more equitable distribution of power in the exchange. By making a message's reception transparent to the sender, some of the anxieties of sending a message off into the void are alleviated, while simultaneously taking pressure off the recipient to fire off a response simply to signal that it has been received. At the same time, it makes reading the message itself a signal – perhaps reinforcing some of the subtle social pressures Vallor identifies as important to shaping communicative virtues, which have previously been neglected. In addition, it preserves boundaries in some of the positive ways discussed earlier.

This illustrates the potential for new technologies to split previously bundled concerns. When we direct our attention to the small ways that communication channels shape exchanges, they can reinforce character traits and values, while suggesting that users themselves may desire tools that maximize the quality of their friendships – and their friends. Perhaps impatience and instrumentality can be discouraged by thoughtful design features, and the popularity of Snapchat suggests that this may be appealing to users.

8. CONCLUSION

I conclude that CMCs can both enhance *and* threaten friendships. The freedom of choice that CMC enables is a double-edged sword. This does not mean that its positive and negative influences cancel each other out, however. The most valuable relationships are those that are choice worthy and so reinforced by the autonomy offered by CMCs, which can also promote quality relationships by screening out those which fail on this front. While most real relationships fall somewhere between the extreme in which all interactions are freely chosen wholeheartedly, and the extreme case of a relationship where every interaction is forced, there is a strong relationship between choice-worthiness of interaction and the quality of a relationship.

To the extent that CMCs pose a moral hazard to relationships, they do so by making it easy for people to respond only to those connections they find worthwhile, and to ignore or opt out of conversations they do not find worth their time and attention. This leads people to be vulnerable in such relationships, as they can only proceed by mutual agreement and mutual engagement; either party's choosing to opt out can signal the end of a relationships. But relationships are better when they are good for each person in the relationship. Users bear the responsibility for reflecting carefully to ensure that their actions reflect their considered evaluations of which relationships, and for deciding which communications, are worthy of response, and when. Vallor is correct to note that users may be tempted to give in to impatience, but this is to some degree counter-balanced by asynchronous communication's reinforcement of patience when it comes to others' projects and schedules.

Furthermore, CMCs can strengthen extant friendships. To the extent that CMCs enhance people's ability to advance their own wellbeing, friendships' quality will be enhanced, at least among those which survive the ability of participants to 'opt out'. This can be clarified by explaining friendships as composite social entities with friends as parts, where the interactions between parts constitute people's composing the friendship. While it is common to claim that friends "share identity", this can give the mistaken impression that friendship involves lowering or removing boundaries between people. The parts/whole account sketched above, by contrast, explains friends as parts of friendships, and clear boundaries or borders between parts can strengthen the integrity of the whole which they compose. Both at the macro level of selecting for good relationships, and the micro level of reinforcing existing friendships, CMCs offer distinct advantages. These advantages can be reinforced with intelligent design features, as illustrated by the example of Snapchat.

9. ACKNOWLEDGMENTS

Thanks to Frances Grodzinsky and Richard Volkman for their feedback on early versions of this paper, and to several anonymous referees at ETHICOMP for their helpful comments.

10. REFERENCES

[1] Aristotle. 1999. *Nicomachean Ethics*. Terence Irwin, Trans. Hackett, Indianapolis, IN.

[2] Badhwar, N.K. 1991. Why it is wrong to be always guided by the best: consequentialism and friendship. *Ethics* 101 (April 1991), 483–504.

[3] boyd, d. 2014. *It's complicated: the social lives of networked teens.* Yale University Press, New Haven, CT.

[4] boyd, d. 2014. Why Snapchat is valuable: it's all about attention. *Apophenia* (March 21, 2014). http://www.zephoria.org/thoughts/archives/2014/03/21/snapchat-attention.html.

[5] Cocking, D. and Kennett, J. 1998. Friendship and the self. *Ethics* 108 (April 1998), 502–527.

[6] Cooper, J.M. 1977. Aristotle on the forms of friendship. *Review of Metaphysics* 30 (June 1977), 619–48.

[7] Dovidio, J. F., Kawakami, K., Gaertner, S. L. (2002) Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology* 82 (Jan 2002), 62-68. DOI=http://dx.doi.org/10.1037/0022-3514.82.1.62

[8] Elder, A. 2013. *Metaphysics of Friendship.* Doctoral Dissertation. University of Connecticut. http://digitalcommons.uconn.edu/dissertations/306

[9] Friedman, M. Autonomy, social disruption, and women. 2005. In *Feminist Theory: A philosophical anthology.* Eds. A. Cudd and R. Andreasen, Eds. Blackwell Publishing, Malden, MA. 339-351.

[10] Johnstone, J. 2007. Technology as empowerment: A capability approach to computer ethics. *Ethics and Information Technology* 9 (March 2007), 73-87. DOI=10.1007/s10676-006-9127-x.

[11] McPherson, M., Smith-Lovin, L., & Cook, J. M. 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* (2001), 415-444.

[12] Plato. 1992. *Republic.* G.M.A. Grube, Trans. Hackett, Indianapolis, IN.

[13] Sherman, N. 1993. The virtues of shared pursuit. *Philosophy and Phenomenological Research* 53 (June 1993), 277-299.

[14] Speer. S.A. and Stokoe, E. Eds.(2011) *Conversation and Gender*, Cambridge University Press, Cambridge, UK.

[15] Turkle, S. 2011. *Alone Together*. Basic Books, New York, NY.

[16] Vallor, S. 2010. Social networking technology and the virtues. *Ethics and Information Technology* 12 (June 2010), 157-170. DOI= 10.1007/s10676-009-9202-1.

[17] Vallor, S. 2012. New social media and the virtues. In *The Good Life in a Technological Age.* P. Brey, A. Briggle, and E. Spence, Eds. Routledge, New York, NY.

[18] Williams, B. 1981. Persons, character, and morality. In *Moral Luck.* Cambridge University Press, Cambridge, UK.

Key Dialectics in Cloud Services

Brandt Dainow
Department of Computer Science, Maynooth University
Kildare, Co. Kildare
Ireland
+353 86 248 2846
brandt.dainow@nuim.ie

ABSTRACT

This paper will identify three central dialectics within cloud services. These constitute defining positions regarding the nature of cloud services in terms of privacy, ethical responsibility, technical architecture and economics. These constitute the main frameworks within which ethical discussions of cloud services occur.

The first dialectic concerns the question of whether it is it essential that personal privacy be reduced in order to deliver personalised cloud services. I shall evaluate the main arguments in favour of the view that it is. To contrast this, I shall review Langheinrich's *Principles of Privacy-Aware Ubiquitous Systems* [24]. This offers a design strategy which maintains functionality while embedding privacy protection into the architecture and operation of cloud services.

The second dialectic is concerned with the degree to which people who design or operate cloud services are ethically responsible for the consequences of the actions of those systems, sometimes known as the "responsibility gap." I shall briefly review two papers which argue that no one is ethically responsible for such software, then contrast them with two papers which make strong arguments for responsibility. I shall show how claims for no responsibility rest on very narrow definitions of responsibility combined with questionable conceptions of technology itself.

The current shape of cloud services is dominated by a tension between open and closed systems. I shall show how this is reflected in architecture, standards and organisational models. I will then examine alternatives to the current state of affairs, including recent developments in support of alternative business models at government level, such as the House of Lords call for the Internet to be treated as a public utility (The Select Committee on Digital Skills, 2015).

CATEGORY

K.4.1 **[Public Policy Issues]**: Ethics

GENERAL TERMS

Design. Human Factors.

KEYWORDS

Cloud services, ethics, privacy, security, privacy by design, personalization, filter bubble

INTRODUCTION

This paper explores dialectics within debates regarding key ethical issues pertaining to cloud services. These issues concern privacy, responsibility for the actions of systems and the development of monopoly service providers. Between them these concerns largely dictate the shape and capabilities of current and future cloud-based services. I shall show how the current state of affairs is dominated by a sense of lack of agency in terms of doing things differently from the current reflexive practice, an assumption that no alternatives to current practice are possible. This paper will attempt to organise the key concerns with cloud services by arranging them into three dialectical axes:

- The nature of the relationship between personal privacy and service provision.
- The degree to which people who build or operate cloud-based services are ethically responsible for the actions or effects of those services.
- The nature of the marketplace for those services.

Since this paper considers cloud services in the broadest sense, it is appropriate to commence with the definition of cloud computing used in this analysis. The *US National Institute of Standards and Technology Special Publication 800-145* defines the essential characteristics of cloud computing as being:

- The ability to provide services whenever desired without human intervention.
- Being available to a wide range of client devices via networking technology.
- The "virtualisation" of computing resources, such that digital operations are not linked to specific servers or locations.
- Scalability – the capability of the systems to scale up or down in response to changes in demand (a necessary corollary of virtualisation).
- Often, but not necessarily, Software as a Service. [29]

Clearly this definition applies to many, if not most, internet systems and digital services, not merely to the virtualisation of server functions previously found in the traditional client-server network. Under this view, Facebook and Google search are both cloud services. I think this is both valid and important - confining discussion of cloud computing to data processing or file storage functions limits discussion to a few contingent uses of a wider system and obscures the essential factors we need to consider.

PRIVACY VERSUS SECURITY

Our first axis is the necessity versus the contingency of reductions to privacy under new digital services. That is to say, there is one body of opinion which holds that the erosion of personal privacy is a necessary and unavoidable consequence of, or precondition for, the delivery of digital services. These positions tend to be a reflexive response within the development community, rarely stated formally, and is a minority view in the literature, as a result of which detailed arguments as to why privacy *must* be reduced to enable cloud services are scarce. However, Lucas Bergkamp's paper, *The Privacy Fallacy* [5] marshals all the arguments in this camp.

Bergkamp argues there should be no privacy protection of any form because preservation of personal privacy is harmful to society in many ways. He provides five main arguments; there is no need for data privacy, data protection reduces individual freedom, personal privacy is contrary to economic growth, EU data legislation is unenforceable and the EU's data protection regimes put it out of step with the rest of the planet. I will now explore each of these in more depth:

Bergkamp argues there is no need for data protection or digital privacy because no one wants it and it serves no purpose. He states there is no evidence anyone has ever been harmed by privacy violations or personalization of services based on personal data. He does not provide any evidence for this and it is contradictory to the reported activity of many data protection authorities. For example, in 2014 the Data Commissioner of Ireland received 2,264 data breach notifications, investigated 960 complaints and launched 162 prosecutions. Half (53%) of complaints involved disclosing personal data inappropriately, such as disclosure of personal financial data to relatives or the listing of email addresses and passwords on public websites [34]. The *Verizon 2014 Data Breach Investigations Report* [46] covers 63,000 data violations across 93 countries in 2014. It highlights financial theft and the cost of dealing with a breach, such as cancelling credit cards, as the main harms to the individual. Other research exists to show harm from less obvious privacy violations. *RT@Iwantprivacy: Widespread Violation of Privacy Settings in the Twitter Social Network* details harm from privacy violations in Twitter when people reuse private tweets in public [28]. *Privacy Violations Using Microtargeted Ads* [21] details harm from privacy violations in Facebook. Privacy violations has also been shown to harm the companies themselves. *How Privacy Flaws Affect Consumer Perception* [2] shows how privacy breaches reduce the chance people will buy from a company, while *Is There a Cost to Privacy Breaches?* [1] shows how privacy violations reduce a company's share price. Studies also exist to show harm from personalization of advertising and news. The research findings of Sweeney's *Discrimination in Online Ad Delivery* [42] reveal how racial stereotyping in ad personalization harms Afro-Americans in many ways, including job prospects and access to financial services. *Bursting Your Filter Bubble* [38] shows harm from news personalization, while the famous Facebook news manipulation study, *Experimental Evidence of Massive-scale Emotional Contagion through Social Networks* [22] shows how personalizing news feeds to contain more negative contents can depress people. Bergkamp's proposition that there has never been any harm from privacy violations or personalization appears to be contradicted by such evidence.

Bergkamp also argues there is no need for data protection because no one wants it. He argues that people don't realise that data protection prevents personalization, but that when they do, they always prefer personalization over data protection. He does not cite any evidence for this. By contrast, Culnan's 1993 study of personalization in shopping, *How Did They Get My Name?* [14] shows that when offered the choice, the people he surveyed preferred privacy over personalisation. More recently, the *2013 Comres Big Brother Watch Survey* [11] polled 10,000 people in nine EU countries to find 75% were concerned about privacy and wanted data protection regulations, while 45% believed they were being harmed by corporate data practices.

Bergkamp also argues there is no need to regulate sale of personal data because companies never sell it. However, there is, in fact, a huge industry in the sale and aggregation of personal data, as the 2014 Federal Trade Commission's investigation into data brokers found [7,17].

Bergkamp argues that personalization results in cheaper prices. However, he does not cite any empirical evidence for this or reasons why it should be so. He cites as evidence a statement made by Fred Cate, Professor of Law at Indiana University, that personalization results in cheaper prices, but this was a statement made to a Congressional committee, not a research finding. Prof. Cate's own list of publications does not include any research into personalization, his speciality is data protection law. Later in the paper Bergkamp states that data protection costs money, and that it is so burdensome and expensive that businesses can only survive by ignoring their legal obligations. One may surmise he believes this is the cause of higher prices to consumers, though he does not explictly say so. However, research like Sweeney's *Discrimination in Online Ad Delivery* [42] shows how personalization actually increases costs to Afro-American consumers in the USA, while Turow's *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World* [44] shows how personalization can reduce or increase prices, depending on whether you are the consumer companies want or not.

Bergkamp also argues that privacy protection increases identity theft because data protection makes it harder to tell if someone really is who they claim to be. He does not cite any evidence for this and it seems counter-intuitive. Given that privacy protection reduces access to the personal data necessary for identity theft, such protection could be presumed to make it harder to commit, so one could argue the exact opposite of Bergkamp in the absence of any research. Bergkamp's position here allies with his arguments elsewhere in his paper that we all need to know as much as possible about each other in order to protect ourselves from one another, and that privacy directly prevents this. He states that one problem with privacy protection is that it allows an individual to control what they disclose to the world. He does not explicitly say this is a bad thing, but it is clearly implied from his usage. Here it is worth noting research showing the reverse, that lack of privacy restricts human freedom. For example, knowledge one is being watched on the internet has been shown to have a chilling effect on what people say [4] and what they search for [25], even when engaging in legal and socially acceptable activity.

Bergkamp claims there is a vast amount of money to be made acquiring and selling personal data, despite his earlier claim that there are no businesses selling it. He provides no evidence for this economic activity, but the claim is supported elsewhere. For example, in 2013 the OECD estimated the personal data of each Facebook user to range from $US40/year to $US400 [33]. Bergkamp claims that this data market alone is sufficient reason to remove privacy protections. However, the mere presence of economic activity does not, in and of itself, mean we should encourage it. There is a vast amount of money to be made in drug smuggling, but no one uses that as an argument for encouraging it.

Bergkamp also states that the EU's data legislation is unenforceable. He says the very concept of personal privacy is too vague to support regulation and that the regulations cannot properly specify what constitutes personal data. Furthermore, he says, each privacy incident must be judged on its own merits. He does not explain how judging a case on its own merits is a problem. Each and every infraction of the law is judged individually, so arguing that this is also the case for privacy issues does not, in and of itself, constitute a sign of poor legislation. Furthermore, it is difficult to imagine what the alternative would be if a regulator or judge was not allowed to consider the specific details of each case they were trying to adjudicate.

As stated earlier, Bergkamp believes that data protection is so onerous that no business can do it properly and survive financially. He claims the only outcome is that data regulations are never enforced. Clearly the many cases of prosecution for privacy violations are not accounted for in this argument. Bergkamp also states that EU data protection legislation is founded on a misunderstanding of how business works, but

does not provide any further details regarding the nature of the misunderstanding or what the reality truly is.

Bergkamp's paper also states privacy protection damages society because it involves the government paternalistically interfering in people's relations with each other in a misguided attempt to stop people hurting each other. Such an argument can also be said of laws against violence and theft, so the logical consequence of such a position is that we should move to a state of complete anarchy. However, Bergkamp does not address this implication. Instead he goes on to state that government's should never restrict any information under any circumstances. Again no reasons are provided to justify this proposition. Such a broad statement can also be used as an argument in favour of making child pornography freely available, so some additional clarification would seem appropriate.

Finally, Bergkamp claims that EU data protection legislation is out of step with the rest of the world. He does not provide any evidence to support this, but he clearly thinks this is a bad thing and grounds for abandoning data protection. This is a questionable claim. The EU's data protection regime was intentionally built to accord with pre-existing OECD guidelines, which were first developed in 1980 [32].

Bergkamp never states that it is technically impossible to maintain privacy while extending cloud services. His arguments are merely that we should not. My position is that there is no necessary and unavoidable relationship between privacy consequences and functionality. One does not *have* to reduce privacy in order to extend services. Rather, it is always a question of choice, either in how the system is constructed or in the type of business model under which it operates, and there are *always* alternatives. It may be that some of those alternatives are more expensive than the privacy-reducing models, or that alternatives are more technically challenging. However, that, in and of itself, is not an argument for the necessity of privacy-reducing models, but rather an argument underpinning a particular business model or software approach.

Currently those who are building cloud-based services most commonly work on the basis that privacy is exchanged for digital services. However, there is also a growing body of those seeking to develop alternatives, in terms of governance or business model or in terms of code. The most notable is the Privacy by Design movement. However, most of the Privacy by Design material is so vague as to be little more than statements of intent. For example, IBM claim to have moved to Privacy by Design by doing nothing more than implementing awareness training and building an internal system for reporting data breaches [35]. Here Langheinrich's paper, *Principles of Privacy-Aware Ubiquitous Systems* [24] stands out as the exception, being a concrete statement of specific technical design principles which genuinely do embed privacy considerations into the technical architecture. Langheinrich's paper shows it is possible to build robust systems which have privacy protection embedded within the design and operation of the system.

It is notable that Langheinrich has practical experience in the design of privacy systems, being one of the authors of the W3C's technical standard, *Platform for Privacy Preferences*, or PPP [13]. The PPP standard enables browsers to hold the user's preferences for what data they will allow a website to gather. The server component of PPP allows the web server to list its own data-gathering practices. PPP then enables the browser to compare the web server's practices with the user's preferences. The system provides for warnings to the user and for compact and rapid communication between client and server of data practices. The system was supported in Microsoft Internet Explorer 6 when it first emerged, but lack of support by website owners means PPP is largely unused today.

The first of Langheinrich's principles is the Principle of Openness, or "Notice." This simply states that no device or service should gather data about someone without telling them. Here he makes reference to PPP as providing a digital vocabulary which could be used to programmatically describe what data is being gathered, for what purpose and by whom. This is paired with the second principle, the Principle of Consent, which encodes the legal necessity for informed consent. A system must allow for someone to opt out of being tracked or recorded, and do so without denying service on a "take it or leave it" basis. Thus, for example, buildings would need to disable tracking for some people and not simply refuse them entry.

The third principle is termed "Anonymity and Pseudonymity." This states that people must have the option to remain anonymous. The issue here is that some services are only possible if they know a user's identity and history. Here Langheinrich introduces pseudonymity. Under this system a person may have a unique identifier of some form, such as a cookie or RFID chip, which anchors the data systems and forms the index key to their personal data history. However, this identifier contains no personally identifiable information and is discardable at any time. Furthermore, such a system permits people to have multiple pseudonymous ID's and so prevent aggregation of disparate activities by data brokers. It is noteworthy that EU data regulations have recently been updated to add the category of pseudonymous identity between personal and anonymous data [48].

Langheinrich's fourth principle of "Proximity and Locality" limits the scope of data collection. Looking to a future in which people have many devices capable of recording their surroundings, the principle of proximity states that these devices can only operate in the proximity of their owner. This prevents people leaving devices to record data unseen, then returning for them later. Of wider application is the principle of locality; devices should not transmit data any further than absolutely necessary to fulfil their functions. For example, Samsung's voice-activated TV's transmit all conversations they hear to Samsung's central servers. Voice commands are interpreted there and the appropriate command then sent back to the TV. All conversation recordings are stored permanently for later analysis [49]. Under Langheinrich's principles the TV would have been designed so that it did not need to involve cloud services. Voice recognition chips have been around for 20 years and could have be used instead.

The fifth principle is the "Need for Security," in which Langheinrich advocates various levels of security depending on the nature of the data. More importantly, he illustrates how the previous principles themselves enhance security. If data is not being transmitted many security problems simply vanish. Similarly, if data is not linked to an identifiable individual, but only to a pseudonymous ID, unauthorised access has less potential for harm.

Langheinrich's final principles are the principles of "Collection and Use Limitation." These state that data collectors should only collect data for a specific purpose and not store it, as Samsung TV does, in case they want to use it in the future. Secondly, they should only collect the data they need in order to fulfil their task and nothing more. Finally, they should only keep data as long as it is necessary for the purpose. While these appear primarily legislative principles, they can be embodied in technical design through the use of the earlier principles. For example, if data is housed in the user's devices in accordance with the principle of locality, then the user can impose usage and storage limitations themselves.

Langheinrich's principles, if implemented, would solve many privacy concerns, enhance security and actually make many applications of ubiquitous and cloud services easier to construct. What they show is that it is perfectly possible to

design cloud services in a manner which enhances both security and privacy at the same time, while permitting all the personalization necessary. They place control of personal data firmly in the hands of the user without compromising technical operations in any way. In fact, their reduced dependence on permanent access to centralised services makes them more robust and reduces the burden of traffic on the internet. These principles are easy to understand and yet produce powerful architectures. They offer a practical and detailed response to the reflexive position that personalized cloud services must reduce privacy. In doing so they provide concrete evidence that it would be possible to move cloud service evolution into a path which fulfils all its potential, yet enhances privacy and security at the same time. Langheinrich's design principles demonstrate that the reduction of privacy in cloud services is a choice, not a necessity.

ETHICAL RESPONSIBILITY

Our second axis is concerned with the degree to which people who design, build or operate cloud services are ethically responsible for the consequences of the actions of those systems. This question does not arise with regard to all cloud services, but only with the rising generation of autonomous services which process personal data in order to deliver personalised services, such as personalised search results, product recommendations and news feeds. In the near future we will see the rise of more intelligent and more life-critical personalised services, most notably with bio-implantation and other medical services [19]. The question is primarily one of who is responsible when such autonomous services make decisions which result in harm, but where these decisions are not the result of faulty design or incorrect data.

The competing positions are that, on the one hand, programmers and operators are not ethically responsible for the actions of autonomous systems, versus a view that they are. It is difficult to argue that the person holding a hammer is not ethically responsible for the consequences of whatever happens when the hammer hits something because the hammer is totally under the control of the user. However, with large industrially-produced complex automated systems, especially those that include some form of AI functionality, arguments emerge in favour of the position that those who build the systems are not ethically responsible for the decisions those systems make. This argument will no doubt be exacerbated the more powerful and the more intelligent and autonomous these systems become. This issue is discussed most frequently with regard to autonomous military systems, whose lethality makes the question of ethical responsibility both stark and urgent. However, the question is just as pertinent for any form of autonomous system, including those cloud-based personalization systems already in operation.

Andreas Matthias' paper, *The Responsibility Gap* [26] offers a fairly straightforward account based on philosophical logic to support the position that programmers are not responsible for the actions of their autonomous systems, while Robert Sparrow's *Killer Robots* [39] presents the same conclusion via an examination of the practicalities of creating and deploying autonomous systems. Both take the position that no one at all is ethically responsible for the actions of autonomous agents.

Sparrow's argument is based on his particular understandings of the terms 'autonomy' and 'responsibility.' My view is that he defines these terms in such a way as to make any contrary conclusion impossible. Early in the paper, he defines autonomy as being free from external causation:

> "Where an agent acts autonomously, then, it is not possible to hold anyone else responsible for its actions. In so far as the agent's actions were its own and stemmed from its own ends, others cannot be held responsible for them. Conversely, if we hold anyone else responsible for the actions of an agent, we must hold that, in relation to those acts at least, they were not autonomous." [39:65–66]

Sparrow does not defend this definition of autonomy. However, once it has been defined this way, it becomes a matter of logical necessity that there is no ethical responsibility by the programmers or controllers. It is also worth noting that Sparrow uses autonomy in an absolute sense, as if the agent were free from all influence except their prior experience. In particular, he does not recognise the environment, the capabilities of the device or its internal structures as having any impact on decision-making. He argues the programmer cannot be responsible because the essence of an autonomous system is that it will make unpredictable decisions. He argues the controller of the system is not responsible because they could not anticipate what it would do any better than the programmer. In both cases, he ignores the fact the system is designed to perform a particular role in a particular environment. A software agent is not free to do just anything, it can only recognise inputs of a type it has been designed for, and has a relatively limited range of actions it can take, and can only operate in a specific type of environment. A share-dealing system cannot walk the dog or assess your exercise regime. The type of decisions an autonomous system may make, and the range of options available to it, are not only predictable, they are the basis upon which it was designed and built - they define it. An autonomous system may make its own decisions, even alter its own programming, but its range of actions and the forms of harm it may commit are knowable in advance in virtue of the type of system it is.

Sparrow does not mention Strawson in his paper, but his conception of moral responsibility has close parallels to Strawson's influential work. Strawson's position is that no one is morally responsible for anything because no one is free from external influence [41], though the details of why are beyond the scope of this paper. Though Sparrow does not say so explicitly, his use of responsibility is clearly that one can only be responsible for specific actions. In Sparrow's view, the design of the system and the decision to use it do not carry any ethical responsibility because neither gives one the ability to predict the specifics of an individual act the system may take.

Sparrow also argues it is not possible to hold the system itself responsible because responsibility necessarily requires punishability which requires suffering. Under his definitions, something can only be morally responsible if it can be punished and something can only be punished if it can suffer. Since software systems cannot be made to suffer, they cannot be punished and so cannot be held responsible for their actions. Note here that we have switched from talk of "being responsible" to talk of "being held responsible." Here we see that Sparrow has conflated the moral state of being responsible with the social status of being eligible for punishment.

Matthias's *The Responsibility Gap* [26] also argues that no one is responsible for the decisions of autonomous software systems. His position also links responsibility to individual acts, holding that one can only be responsible if one can know the internal state of the system *and* has control of each act it takes, at least to the degree where one could prevent it. Under this analysis a programmer has no responsibility for the actions of a system once the owner takes control. The owner is not responsible because they cannot know the internal state of the system. Matthias spends some time examining different types of AI learning, showing how each makes their internal state unknowable in different ways, but the differences do not affect his final conclusion.

The narrow understanding of responsibility seen in Matthias and Sparrow is the foundation on which their arguments rest. In

contrast, Miller's *Collective Responsibility and Information and Communication Technology* [30] confronts this issue by arguing there are different types of responsibility. In addition to the responsibility for individual acts which Matthias and Sparrow focus on, Miller points out we also recognise one can have "structural" responsibility by creating the conditions which made the act possible or by ordering others to take actions which eventually led to the act. Under Miller's analysis both programmers and controllers of autonomous systems take structural responsibility for every act taken by these systems. Miller then goes deeper, investigating the concept of collective responsibility. He argues that to the degree that individuals contribute something to the shape and operation of an autonomous system, so they share in responsibility for its actions. Here he acknowledges the existence of corporate responsibility, but argues that it does not provide a moral shield for the individual workers, whose individual contributions to a system's operation convey a share in collective responsibility for its actions.

Miller's approach is a step towards recognition that software agents exist within the wider context of human activity. This broader perspective is fully achieved in *Software Agents, Anticipatory Ethics, and Accountability* by Johnson et. al. [20]. Reiterating the perspective that technology is socially situated, this paper argues that the concept of any digital service as an autonomous agent is merely metaphorical; that no such system can be autonomous in the sense we apply autonomy to humans in moral debates. As such, the use of the metaphor is justifiable only by its utility. Johnson et. al. criticise the concept of any software as an autonomous agent on the grounds it generates just these ethical problems. Instead, Johnson et. al. argue we should recognise autonomous systems as elements within a larger socio-technical system, made by people and used by people for human purposes. Under this view autonomous systems are not independent entities hermetically sealed from their environments, but systems which can only be understood by reference to the context of their use. Johnson et. al. make implied use of the different forms of responsibility seen in Miller, but do not elucidate the differences. Instead they focus on the arbitrariness of delimiting technical artefacts. They deny that autonomous software agents are different in kind from any other form of automated or semi-automated device, being merely more complicated. Autonomous systems are thus merely, like hammers, extensions of human will and intent. Under this arrangement, ethical responsibility for their actions is not in any way changed by the mere fact of their complexity.

OPEN VERSUS CLOSED

Our final dialectic concerns the form and marketplace of cloud services. Here the dominating dialectic is that of open versus closed systems and open versus closed organisational contexts for such systems.

The scene for this debate is best set Eben Moglen in his presentation, *Freedom in the Cloud*, delivered to the Internet Society in 2010 [31]. Moglen argues that the internet was originally designed as a non-hierarchical peer-to-peer network. However, under the influence of the architectural model of client-server networking, the services which evolved used a smart-server-dumb-client model, in which both algorithms and data were centralised. Moglen maintains that cloud architecture works on this thin client - fat server model and does not represent a new computing architecture, merely the virtualisation of some server operations within this traditional model. These servers maintain activity logs. These logs can be mined for behavioural data. Marketing companies learned they could mine these logs to understand, predict and influence user behaviour in order to sell advertising. Moglen contends that as the perceived value of this information grew, it spurred the development of a secondary internet infrastructure of tracking services designed to add to the growing database of what we now call "user profiles."

Thus Moglen describes how an architecture which concentrates processing power and data at centralised locations promotes a concentration of both technical proficiency and economic power, while also promoting a top-down hierarchical organisational model and, in a global internet, the development of a limited number of very large monopoly service providers. This has, he argues, produced an extreme power dichotomy between those who own the services and those who use them. The business model which has come to dominate the internet is that of delivering services in exchange for spying on the users all the time. Moglen describes this state of affairs as undesirable for two reasons. Firstly, the price is too high and the services are not worth the loss of privacy. Secondly, the lack of alternative models for access to the same services makes this unfair arrangement unavoidable. He argues that we need an alternative architecture in which the data about us stored on centralised servers is instead housed in devices we own and carry with us. We can then control who accesses this data and how. He argues that this is possible with current technology.

There is an additional element of concern within Moglen's model which he hints at but does not explore. The combination of architecture and business model he describes has produced "walled gardens." These are silos of private technology and proprietary data formats which are not compatible with, or accessible by, other systems or organisations. The patent system combines with a capitalist marketplace to financially reward such behaviour. If I am the sole owner of a system everyone wants to use, I can make money. If I create a system which I give away, I do not benefit. What I therefore need to do is lock everyone into my technology, and then I will "lock in" the market [12].

The effect of this is to lock data and services into a single monopoly provider. The provider becomes the gatekeeper over the knowledge of what they do and how they do it. Users cannot migrate to a competitor without significant effort and loss. For example, if you close your account with Amazon, they will remove all the books from your Kindle [50]. You cannot therefore switch to an alternative, such as Adobe Digital Editions, without re-purchasing your entire digital library. Different legal regimes permit different levels of access inside these walled gardens, but in no case does a society have full knowledge or any substantive control. Such a system has no interest in open standards, interoperability, or a free flow of information. This lack of interoperability and open standards was why the internet and HTML were not developed by commercial enterprises. Early pre-cursors of the web tried the same walled garden approach, including America Online, CompuServe and Lotus Notes. It was only when Tim Berners-Lee gave HTML away that we broke free of this limiting system and gained the web. Berners-Lee gave it away because he saw things in exactly this way and believed that if he patented or sold HTML, it would become just another walled garden [6].

However, as companies have developed services which sit atop these communally-owned standards, so they have developed further proprietary systems. The final result is that companies have built a new layer of walled gardens and data silos on top of the open platform which is the internet [16]. The scale of the internet user base combines with a shared service delivery infrastructure to enable the rise of extremely large global monopolies, such Google, Amazon and Facebook. The result is that cloud services are portioned out amongst a limited number of very large hierarchical organisations, each of which hides its use of data from public scrutiny and uses its monopoly position and ownership of personal data as a competitive advantage [8,15,27]. The net effect is that people are locked to service

providers like serfs to their lord. However, unlike in the Middle Ages, there is no competing lord to flee to if you are unhappy with your lot. This power is a concern to many. Some argue, for example, that Amazon's potential to control what books are available makes it a political institution as well as an economic one [10], while Google has consistently ranked as one of the biggest spenders on political lobbying in Washington, D.C. since 2012 [18].

Opposing this state of affairs are a disparate range of alternatives, such as Moglen and his concept of a personal server. Each alternative tends to focus on one aspect of this system, such as technical architecture or business model. Technically, the existence of the internet is based on open standards, such as TCP and IP [40], so alternatives have always been available on a technical level. Here we have the open source activists, such as the Free Software Foundation and the IETF. In addition, we have less obvious alternative architectures based on peer-to-peer (as opposed to client-server) models, such as the BOINC platform for community computing [3] and the BitTorrent protocol [36]. Standards like XML [47] and RDF [37] provide a means of breaking open walled gardens through data exchange, while people such as Chris Marsden in the UK or Robert McChesney in the USA have developed the rationale for breaking down these proprietary data silos.

McChesney argues that the development of monopolies and cartels has so dominated the internet that there has been little economic benefit for the rest of society. He argues that there is so little competition at the point of delivery that service providers constitute a cartel which should be forced into competition with not-for-profit public alternatives. He calls for the monopolistic corporations dominating important services, like Facebook and Google, to be broken into smaller competing units and subject to much more stringent and detailed state control. McChesney's argument is that the size of these corporations is so great they pose a threat to democracy itself through their power to lobby politicians, dominate online debate and skew economic development [27].

Concern over monopoly domination is addressed in a different manner by Brown and Marsden in a number of publications. Instead of seeking a solution by changing the economic structure, they focus on the proprietary data structures which form the foundation of such domination. In addition to rights such as the right to have one's records deleted, they argue for the right to move such data to an alternative provider of the same service [9]. They cite similar historical examples in which Microsoft, IBM and Intel have been forced into making their systems interoperable with competitors, mainly through antitrust approaches in the USA and EU [8]. They argue that state intervention to break up these monopolies is not practical in a world dominated by competing national legislative regimes. Instead, they argue that merely providing users the ability to switch to alternatives would be sufficient. They believe that this would stimulate the development of service providers offering a range of alternative models [10].

The approach of treating monopoly service providers as public utilities is gaining ground in government circles. Recently the UK House of Lords called for the internet to be treated like a public utility rather than a market place of optional luxuries [43]. International bodies, such as the EU and UNESCO, have started calling for wider civic involvement in determining how services are provided [23,45] and for the development of alternative service provision models. For example, the outgoing EU Vice President, Neelie Kroes, stated in November 2014:

> "Why should we have to give up our privacy for a "free" service if we prefer to pay for that same service with cash and keep our privacy?" [23]

CONCLUSIONS - AGENCY

While these three dialectics focus on different issues, the poles of each axis rest on competing perspectives on the possibility of agency. Those who accept things as they are now do not see a possibility for agency, while their opponents do. On the first axis we have those who hold that preservation of privacy and delivery of service are necessarily in opposition. Here they are holding that there is no possibility of agency in the relationship between privacy and service design. To the contrary, we have seen how Langheinrich's design principles show multiple opportunities to intervene in the ways which deliver services while also maintaining privacy. In our second axis of ethical responsibility, the position of there being no ethical connection between the creator of an autonomous system and that system's effects is also a position of there being no agency. Here lack of agency pertains not to the nature of the system, but to the consequences of the system's actions. Under this view, once an autonomous system is activated, human agency ceases. However, as we have seen, preserving a lack of responsibility requires limiting the conception of where agency lies. Responsibility has to be defined in a very constricted manner which focuses on the making of each individual decision and denies the influence of any context. Instead, services are treated as independent of any human agency, in terms of their design, their environment, their purpose, how they are used and who benefits. By contrast, once autonomous services are contextualised within a field of human practice, human agency becomes apparent throughout the construction and operation of such systems and human ethical responsibility becomes self-evident. Finally, in our third axis of service architecture and business model, we see a historical lack of agency in the development of the broader internet culture. Here the client-server structure was accepted reflexively by developers and users, along with the most obvious reflections of this in organisational and economic models.

In all three debates we see one pole in each dialectic disempowering itself, primarily because it simply fails to recognise that there is a choice and that agency, the power to act differently, exists. The conclusion which emerges from this is that a key step to improving the current ethical status of cloud services is inculcating in programmers and leaders that they possess agency, bringing them to recognise there are alternatives and that they have the power to explore them.

REFERENCES

[1] Alessandro Acquisti, Allan Friedman, and Rahul Telang. 2006. Is there a cost to privacy breaches? An event study. *Proceedings of the Twenty-Seventh International Conference on Information Systems (citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.2942&rep=rep1&type=pdf*.

[2] Sadia Afroz, Aylin Caliskan Islam, Jordan Santell, Aaron Chapin, and Rachel Greenstadt. 2013. How Privacy Flaws Affect Consumer Perception. IEEE, 10–17. http://doi.org/10.1109/STAST.2013.13

[3] D.P. Anderson and G. Fedak. 2006. The Computational and Storage Potential of Volunteer Computing. *CCGRID '06 Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid*, IEEE Computer Society, 73–80. http://doi.org/10.1109/CCGRID.2006.101

[4] Katy Glenn Bass. 2013. *Chilling Effects: NSA Surveillance Drives U.S. Writers to Self-Censor*. PEN American Center. Retrieved from http://www.pen.org/sites/default/files/Chilling%20Effects_PEN%20American.pdf

[5] Lucas Bergkamp. 2002. The Privacy Fallacy. *Computer Law & Security Report* 18, 1, 31–47.

[6] Tim Berners-Lee. 1999. *Weaving the Web*. HarperCollins, New York, N.Y.

[7] Nathan Brooks. 2005. *Data Brokers: Background and Industry Overview*. Congressional Research Service, The Library of Congress, Washington, D.C.

[8] Ian Brown and Christopher T. Marsden. 2013. *Regulating code: good governance and better regulation in the information age*. The MIT Press, Cambridge, Mass.

[9] Ian Brown and Christopher T. Marsden. 2013. Regulating Code: Towards Prosumer Law? *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2224263

[10] Ian Brown and Christopher T. Marsden. 2013. Interoperability as a standard-based ICT competition remedy. *8th International Conference on Standardization and Innovation in Information Technology 2013*, IEEE, 1–8. http://doi.org/10.1109/SIIT.2013.6774570

[11] Comres. 2013. *Big Brother Watch Online Survey*. Comres, London.

[12] Eric P. Crampton and Donald J. Boudreaux. 2003. Does Cyberspace Need Antitrust? In *Who Rules the Net?*, Adam D Thierer and Clyde Wayne Jr. Crews (eds.). Cato Institute, Washington, D.C.

[13] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. W3C. Retrieved from http://www.w3.org/TR/P3P/

[14] Mary J. Culnan. 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly* 17, 3, pp. 341–363.

[15] James Curran, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. Routledge, London.

[16] Tony Dyhouse. 2010. Addressing the silo mentality. *Infosecurity* 7, 2, 43. http://doi.org/10.1016/S1754-4548(10)70043-7

[17] Federal Trade Commission. 2014. *Data Brokers: A Call for Transparency*. Federal Trade Commission.

[18] Tom Hamburger and Matea Gold. 2014. Google, once disdainful of lobbying, now a master of Washington influence. *The Washington Post*. Retrieved June 23, 2015 from http://www.washingtonpost.com/politics/how-google-is-transforming-power-and-politicsgoogle-once-disdainful-of-lobbying-now-a-master-of-washington-influence/2014/04/12/51648b92-b4d3-11e3-8cb6-284052554d74_story.html

[19] Veikko Ikonen, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. *D.1.2. Emerging Technologies Report*. ETICA Project.

[20] Deborah G. Johnson. 2011. Software Agents, Anticipatory Ethics, and Accountability. In *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight*, Gary E. Marchant, Braden R. Allenby and Joseph R. Herkert (eds.). Springer Netherlands, Dordrecht, 61–76. Retrieved February 17, 2015 from http://www.springerlink.com/index/10.1007/978-94-007-1356-7_5

[21] Aleksandra Korolova. 2011. Privacy Violations Using Microtargeted Ads: A Case Study. *Journal of Privacy and Confidentiality* 3, 1.

[22] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. 2014. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111, 24, 8788–8790. http://doi.org/10.1073/pnas.1320040111

[23] Neelie Kroes and Carl-Christian Buhr. 2014. Human society in a digital world. *Digital Minds for a New Europe*. Retrieved January 27, 2015 from https://ec.europa.eu/commission_2010-2014/kroes/en/content/human-society-digital-world-neelie-kroes-and-carl-christian-buhr

[24] Marc Langheinrich. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *Proceedings of the Third International Conference on Ubiquitous Computing*, Springer-Verlag, 273–291. Retrieved from http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf

[25] Alex Marthews and Catherine Tucker. 2014. Government Surveillance and Internet Search Behavior. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2412564

[26] Andreas Matthias. 2004. The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 6, 3, 175–183. http://doi.org/10.1007/s10676-004-3422-1

[27] Robert W. McChesney. 2014. Be Realistic, Demand the Impossible: Three Radically Democratic Internet Policies. *Critical Studies in Media Communication* 31, 2, 92–99. http://doi.org/10.1080/15295036.2014.913806

[28] Brendan Meeder, Jennifer Tam, Patrick Gage Kelley, and Lorrie Faith Cranor. 2010. RT@ IWantPrivacy: Widespread violation of privacy settings in the Twitter social network. *Proceedings of the Web*, 1–12.

[29] Peter Mell and Timothy Grance. 2011. *The NIST Definition of Cloud Computing*. National Institude for Standards & Technology.

[30] Seumas Miller. 2008. Collective Responsibility and Information and Communication Technology. In *Information Technology and Moral Philosophy*, Jeroen van den Hoven and John Weckert (eds.). Cambridge University Press, Cambridge; New York, 226–250.

[31] Eben Moglen. 2010. Freedom in the Cloud. Retrieved from http://www.softwarefreedom.org/events/2010/isoc-ny/FreedomInTheCloud-transcript.html

[32] OECD. 2013. *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. OECD. Retrieved from http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm

[33] OECD. 2013. *Exploring the Economics of Personal Data*. Retrieved June 23, 2015 from http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en

[34] Office of the Data Protection Commissioner. 2015. *Annual Report of the Data Protection Commissioner of Ireland 2014*. Data Protection Commission of Ireland.

[35] Office of the Information & Privacy Commissioner of Ontario and IBM. 2011. *Privacy by Design: From Theory to Practice*. Office of the Information & Privacy Commissioner of Ontario, Ontario.

[36] Johan Pouwelse, Paweł Garbacki, Dick Epema, and Henk Sips. 2005. The Bittorrent P2P File-Sharing System: Measurements and Analysis. In *Peer-to-Peer Systems IV*, Miguel Castro and Robbert van Renesse (eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 205–216. Retrieved June 23, 2015 from http://link.springer.com/10.1007/11558989_19

[37] RDF Working Group. 2015. RDF - Semantic Web Standards. *RDF - Semantic Web Standards*. Retrieved June 23, 2015 from http://www.w3.org/RDF/

[38] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: strategies for promoting diverse exposure. *Proceedings of the 2013 conference on computer supported cooperative work companion*, ACM Press, 95. http://doi.org/10.1145/2441955.2441981

[39] Robert Sparrow. 2007. Killer Robots. *Journal of Applied Philosophy* 24, 1, 62–77. http://doi.org/10.1111/j.1468-5930.2007.00346.x

[40] W. Richard Stevens and Gary R. Wright. 1994. *TCP/IP Illustrated Volume 1: The Protocols*. Addison-Wesley, Reading, Mass.

[41] Galen Strawson. 1994. The Impossibility of Moral Responsibility. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 75, 1/2, 5–24.

[42] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2208240

[43] The Select Committee on Digital Skills. 2015. *Make or Break: The Digital Future*. The Authority of the House of Lords, UK. Retrieved from http://www.publications.parliament.uk/pa/ld201415/ldselect/lddigital/111/111.pdf

[44] Joseph Turow. 2011. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World*. Yale University Press, New Haven.

[45] UNESCO Secretariat. 2014. *Internet Universality*. United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/internet_universality_en.pdf

[46] Verizon. 2014. *2014 Data Breach Investigations Report*. Verizon Enterprise Solutions.

[47] XML Working Group. 2015. XML Core Working Group Public Page. *XML Core Working Group Public Page*. Retrieved from http://www.w3.org/XML/Core/#Publications

[48] 2015. *Future-proofing Privacy*. Hogan Lovells, London.

[49] Privacy | Samsung UK. Retrieved June 18, 2015 from http://www.samsung.com/uk/info/privacy-SmartTV.html

[50] Amazon.com Help: Kindle Terms of Use. Retrieved June 22, 2015 from http://www.amazon.com/gp/help/customer/display.html?nodeId=200506200

The Creation of Facts in the Cloud – a fiction in the making

Don Gotterbarn
ACM Committee on Professional
Ethics
Gotterbarn@acm.org

ABSTRACT
Like most significant changes in technology, Cloud Computing and Big Data along with their associated analytic techniques are claimed to provide us with new insights unattainable by any previous knowledge techniques. It is believed that the quantity of virtual data now available requires new knowledge production strategies. Although they have yielded significant results, there are problems with advocated processes and resulting facts. The primary process treats "pattern recognition" as a final result rather than using "pattern recognition" to lead to yet to be tested testable hypotheses. In data analytics, the discovery of a pattern is treated as knowledge rather than going further to understand the possible causes of those patterns. When this is used as the primary approach to knowledge acquisition unjustified inferences are made - "fact generation". These pseudo-facts are used to generate new pseudo-facts as those initial inferences are fed back into analytic engines as established facts. The approach of generating "facts from data analytics" is introducing highly risky scenarios where "fiction becomes fact" very quickly. These "facts" are then given elevated epistemic status and get used in decision making. This, misleading approach is inconsistent with the moral duty of computing professionals embodied in their Codes of Ethics. There are some ways to mitigate the problems generated by this single path approach to knowledge generation.

Categories and Subject Descriptors
H.3.0 [**Information Storage and Retrieval**]: Systems and software – *question answering.*

K.4.0 [**Computers and Society**]: General

K 4.1 [**Public Policy Issues**] – *Ethics*

K.7.0 [**The Computing Profession**]: Professional Ethics – *Codes of Ethics.*

General Terms
Reliability, Experimentation, Security, Human Factors, Standardization, Theory, Legal Aspects, Verification.

Keywords
Data Integrity, Big Data, Data Misuse, Professional Responsibility

1. INTRODUCTION
In the new virtual society living in the Cloud, big data and data analytics use an epistemological approach which creates a virtual description of the world that may be dangerously inconsistent with the real world. The form of logic used in these systems are harmful both to society and its citizens; fiction becomes 'fact,' half-truths become axioms.

Professional computing societies have been concerned with accurate information and concerned with the computing professional's obligations to preserve and protect information. The ACM Code of Ethics from 1992 states that computer professionals should:

1. take "precautions to ensure the accuracy of data,"
2. protect "it from unauthorized access or
3. accidental disclosure to inappropriate individuals."
4. establish procedures "to allow individuals to review their records and correct inaccuracies."
5. define retention and disposal periods for information. [1, Imperative 1.7]

2. DROWNING IN A SEA OF INFORMATION
2.1 Information Generation
The enormous increase in the quantity of data being produced is both a problem and an opportunity. The amount of data now available and being generated can be used to further our knowledge and the ability to share that information over the Internet/Cloud helps produce new insights. However the enormity of the amount of data is no longer manageable without the help of computerized tools. We need the tools to capture and mange this new data.

2.2 Tools used to stay afloat
2.2.1 Professional Concerns
As computing professionals we need to be sure these tools and the methods for using them do not mislead us. Sometimes rapid enthusiastic unstructured adoption of new technologies adoption of new technologies has unanticipated negative consequences and does not help in way intended. In California, under Megan's Law, a publically accessible database of sex-offenders was established to "allow residents peace of mind" knowing whether predators lived in their neighborhoods. In some neighborhoods in Los Angeles there were no areas that were free of sex offenders [21] and the database may have been used to target people for slaying. [21]. any approaches to use and manage this flood of data must be consistent with professional moral obligations.

2.2.2 Inference making and swimming (floundering) in a sea of data

Faced with large quantities of information we make inferences based on selected pieces of that information and our past experiences. Tomes have been written describing the types of errors we make drawing conclusions based on insufficient data or looking at the wrong data. Given the amount of data available in the Cloud, insufficient data is not a problem. Unfortunately given the amount of data what data is selected may still be a problem. Using Internet data to make inferences is not a new process. We have learned to make discriminations about the quality of Internet data. Data also has a provenance- a context which may include its origin, and epistemic quality. Data without such provenance sometimes leads to unjustified inferences as when an insurance company denied a 'non-drinker' discount on the automobile insurance policy to someone because they had visited several websites which sold liquor. [11]

Sometimes misleading data is used to make inferences and generate new data. In a small town in the USA a bank manager cross checked a list of cancer at the local hospital with a list of people with loans from his bank. Concerned about repayment by cancer payments, he recalled the outstanding loans wherever he found a match [11]. His decision changed their digital financial histories to include a 'recalled loan'. This 'fact' would lead others to conclude that the patients were financially unstable and potentially damaging their further financial transactions. Knowing the provenance of this information would show the error of this inference.

The current generation of data and data about data is an opportunity to further knowledge but the sheer quantity of the data makes it difficult to determine which data we can rely on in decision making, which data will prevent us from drowning drawing the wrong conclusions.

2.2.3 Emerging from a sea of data- Police Intelligence AND investigation

Recently I had the privilege of working with police criminologists from several countries and learning how they gather and handle information; learning about a significant distinction between police 'intelligence' and police 'investigation'.

Although the precise definitions of "intelligence" and "investigation" differed slightly from nation to nation, there was a common operational definition. During 'intelligence' the analyst gathers, evaluates and relates significant discrete pieces of data and develops hunches about things in need further investigation. Intelligence analysts have to sift through massive amounts of information whose relevance may be unknown. Because of the adversarial nature of a police data gathering, the information may be incomplete, inaccurate, or hidden. Criminal Intelligence analysis looks primarily at data related to unlawful events to identify patterns between crimes in different places. The police analyst does not reach conclusions about potential or actual criminal behavior. Their insights are then used in criminal investigations.

An investigator forms some testable hypothesis from the criminal intelligence information and gathers confirming and disconfirming information/evidence that could be used as evidence in a court of law as a proof. To qualify to be used as "evidence", the data needs other characteristics such as a reliable source, a provenance showing the acceptability of the context from which it was derived and how it was derived. Standards to determine the usability of information differ slightly between national police agencies. Investigation terminates with decisions; who committed the act, a secure chain of evidence, and a decision to prosecute or not; a decision about whether the hypothesis has been proven (beyond a reasonable doubt). This 'intelligence' – 'investigation' interaction resembles the scientific method.

Because of the complex nature of the data considered in police intelligence, criminal intelligence frequently stores the information in computer systems to later be searched to help identify possible meaningful patterns. Technology has been used to help improve these methods.

3. BETTER LIFE JACKETS

3.1 Data gathering and production

Each advance in path of technology bring new ways to gather and use information. Internet use tracking tools such as Double Click, AdSense, and AdWords which communicate web interactions to Double Click enabled sites and communicate all the data in your double.click cookie to external site.

The tools which recorded your Internet interactions have been supplemented by a new variety of tools which support data analytics. Data analytics includes a series of tools and methods to the make of inferences- generate new information - from these interactions and other data available about you. These inferences are based on correlations which may be misleading so if unexamined these tools and methods may be misinformation propagation devices.

4. THE NATURE OF THE RIVER – THE CLOUD AND BIG DATA

4.1 A Life Raft- The Cloud

Estimates of the number of digital files is quite large and doubling every two years; "the digital universe is something to behold — 1.8 trillion gigabytes in 500 quadrillion "files" — and more than doubling every two years." [10] Storing this quantity of data is beyond the financial and space resources of most corporations. These resource limitations encouraged the development of "Cloud Computing" as either a centralized public or private places to store the data which can also be rapidly access by a diverse group of users.

The Cloud needs to store a diversity of information types from a variety of sources; social media, Internet transactions, public records, etc. Cloud service providers also offer a variety of programs to help analyze and access the data that is stored. As with any highly marketed product there are a numerous

definitions of it. Cloud services are offered by large companies including Amazon, Google, and Microsoft. They offer services to manage the increased volume velocity and variety of data.

Cloud computing faces common computing problems of maintaining security, data integrity and privacy. Cloud service providers are constantly challenged and constantly work on technical counter measures for attacks attempting to gain unauthorized access to their files. But there are also problems unique to the Cloud, such as and a lack of standardization about storage and data management between Cloud service providers although they my use similar technologies. The competitive environment encourages providers to develop their services which distinguish them from other providers.

4.2 Managing data in the Cloud

The volume, variety and complexity of data now being stored is new. Cloud storage has data on the number and size of raindrops in rainforests, and the tweets of every teenager in the USA. The volume is increasing at incomprehensible rates. Every 2 days we create 5 exabytes of information; what it "took from the dawn of civilization to 2003 to create" [14]

Big Data analytic tools and methods were developed to examine this raw instructed data in order to reveal important interrelationships that were previously difficult or impossible to determine. The volume of the data and the speed at which it is produced outpace humans' ability cope with the heterogeneity of the data and to put it into standardized formats. We can make inferences because of known relationships but big data coming from new sources may not be subject to past insight. The data includes algorithms to assemble shadow data – traces of information created by our daily activity- into meaningful items. Our data shadow is used conclusions about us.

In addition to the enormous technical problems creating and managing this technology, there are significant ethical problems that we have a professional responsibility to address.

4.2.1 Digital Shadows- faded connections to reality

Individual create more information about themselves as they move through the digital environment — writing documents, taking pictures, downloading music, etc. — than is being created about them in the digital universe. Your shadow is made up both public and private information [10] and, unfortunately, "pseudo-facts" If you are reading this electronically remote computers are creating models of who you are looking at current and yet-to be-discovered intersections of data. The construction of these digital shadows is based on incomplete and heterogeneous data.

Digital shadows from heterogeneous sources for varying provenance are added to my digital identity. In ordinary life when one makes some inferences about me and tells those hunches to someone else those hunches are be relegated to the category of rumor unless confirmed by other things. When the computer makes similar inferences they become part of my digital identity. Without my knowledge, my *digital persona* has changed. Things like this occur. A programmer, assigned to determine which programming languages are most in use, counts the frequency of the programming languages required for jobs advertised on the web. Her Internet transactions have been tracked and a pattern analysis tool attributes low job satisfaction to her because she spends time looking at job advertisements. Because this was a computer "Insight" it is given inflated epistemic status.

4.3 Hyperbolic claims- Inflated epistemic status

Unfortunately, the use of new technologies and the availability of vast amounts of heterogeneous information - Big Data- facilitated a new emphasis on one particular epistemological approach to knowledge acquisition – pattern identification and data analytics -which has led to an unjustified confidence in the truth of claims which are at best conjecture. [22] Because of the quantity and variety of data used in these conjectures they are unjustly elevated to highly probable or even axiomatic level of trust. The concern is that mistakes made with big data are especially problematic because of the epistemic status given prompted or derived digital "facts". We all make mistakes but these mistakes are more dangerous because of their elevated epistemic status.

5. REASON

5.1 The Reasoner

A new type of computing professional, a data scientist, is emerging to make sense out of large streams of digital information flowing into organizations. . They use big data to model complex business problems, discovering business insights, and identifying opportunities. Data analytics has been used to analyze Google searches to predict flu outbreaks (Google Flu Trends –GFT), phone records to anticipate terrorist activity, and shipping data to detect smuggling activity.

On the positive side, the claim is that they with massive amounts of data and modern computing they can apply data analytics to solve almost any problem and predict events. "In the next 20 years we will be able to predict events and make decisions about such things as how to teach, romantic relationships and who is a criminal and where the next crime will occur" [25] , This kind of praise is common in the literature.[6,25]

5.2 Reasoning constrained by the nature of the data

Given this claimed potential impact we need to be clear about the limits of reasoning with big data. What can be concluded with confidence and what requires further investigation.

How do data analytics – heuristics to derive conclusion from big data- stand in relation to the scientific method mentioned above? Big Data has many defining characteristics. One description characterizes big data as "comprehensive", messy", and "the triumph of correlations" [20]. Using comprehensive big data significantly reduces the problem of not having enough data to work with. In theory, there is no longer a problem of only having the irrelevant data or the wrong data since you have all the data. Messy big data, unlike exact carefully measured scientific data, includes all types of data from the real world and assembled data from digital shadows. Because the data's provenance is not clear

and varies in quality, and is distributed in different formats across countless servers [20]. "Exactitude" is not expected and approximations are satisfactory. "With big data, we'll often be satisfied with a sense of general 4 direction rather than knowing a phenomenon down to the inch, the penny, the atom" [20] It is claimed that correlations provide the answer.

It has also been claimed that big data science will be judged on its containing one or more of "three criteria: Does it provide more useful information? Does it improve the fidelity of the information? Does it improve the timeliness of the response?" [10] But if we are to guide our lives by the conclusions of data analytics then fidelity is a not optional. There are technical ways to protect the data so it does not change- data integrity. However the question is data fidelity- is it an accurate description of the facts that can be used to make decisions and used to draw other conclusions. It may be effective in uncovering high correlations? Is that enough?

Others have maintained that "although big data is very good at detecting correlations, especially subtle correlations that an analysis of smaller data sets might miss, it never tells us which correlations are meaningful." [19]

5.3 Ethics and the data

Data scientist are computer professionals bound by the information standards in the code of ethics. From the analysis above, big data science and data analytics are not consistent with:

"1 take "precautions to ensure the accuracy of data," because of digital shadows a problem with

"4 establish procedures "to allow individuals to review their records and correct inaccuracies" and because we don't even know what is collected it is difficult to

"5 define retention and disposal periods for information."

Are there other problems with knowledge development in big data science? How does data analytics measure up to the police model of knowledge acquisition described above?

6. BIG DATA SCIENCE AND SCIENCE

6.1 Not Science

Police knowledge acquisition is similar to a scientific model where intelligence identifies patterns which can be used to develop theories (hypotheses) and these hypothesis are checked against the data (repeated experiments). Facts are the event data confirming the hypothesis. Big data science identifies patterns without seeking understanding. 'Science' is reduced to finding interesting patterns without finding explanation for these patterns. The "insights" (pseudo-facts) are used as input data fed back into the system to generate further correlations.

Big data must rely on unstructured data. Even if there was enough manpower to structure the data into common formats and filter out misleading content, the heterogeneous big data is only partial, variable in nature, and comes in many formats. There are no standards for assembling it into a single verifiable thread. There is also a strong business disincentive for standardization. Unique services facilitated in part by unique data structures give one provider competitive advantage over another.

6.2 Dangers of trusting inflated epistemic value

Predictive analytics currently based on structured and unstructured data are used in sales and to determine customer loyalty. Some are interested in high correlations being used to take proactive behavior preventing predicted actions, at least on movie has been based on using this predictive (pre-knowledge) insights to lead to people being detained before they commit a crime. The pre-crime analysis will predict who will commit crimes. The inflated epistemic value of big data is ethically dangerous.

Data science is different from the traditional scientific method used by the police. The optimism of data science/ e science also leads to several reasoning mistakes weakening the fidelity of their "insights.

6.3 Data quantity versus data quality- the big guy isn't always right

Data quality degrades exponentially. Initially the data is not collected in a structured way but is the result of irrelevant social functions. Leonelli [18] describes the journey of data which is first decontextualized to fit into other database, then re-contextualized in the received database to include metadata about quality and reliability of the information. This requires a large manual effort which only can come from companies of labs with large resources generating "selective data sets" reducing the possibility of truly comprehensive data. This creates an imbalance in the in the types and sources of data assembled. But these dataset will shape future research. This biased data will be fed back into the system creating an error- amplifier where the biased data will yield higher correlations.

It is misleading to argue that big data science is comprehensive and as such intrinsically unbiased rather than helpful in shaping scientific as well as public perceptions of the features, opportunities and dangers associated with data-intensive research. [17]

Why would few claim that "all along, as the mounds of data continue to bury us, we make little progress in the only thing that matters: doing something useful with data. That's because we've been going about it all wrong"? [9]

To understand what he might mean by "…we've been going about it all wrong." We need to look at way in which "relevant" data is selected for from a big data set and some associated logic problems.

There are two types of problem when we are talking about high correlations and insight discovered by big data science/ The first has do with the way in which we select data and the second with conclusion from correlations found in big data.

7. DATA QUALITY AND THE LADDER OF INFERENCE

When faced with massive and complex data, individuals focus on different pieces of data and add meaning to their selected data. They then make assumptions which form beliefs they act on. These actions have impact on the generation of new data. This is similar to Leonelli's problem of biased data collections (above) caused by economic factors but in this case the bias is generated by individuals and their belief systems affecting data selection.

This "ladder of inference" [3] model, like one version of the scientific method, starts with the collection and examination of data but then the 'ladder of inference' and the scientific method diverge. The scientific method moves to formulate hypothesis or explanations of the data and those explanations are tested by making a predictions. These predictions are organized into tests to check the hypothesis with repeatable experiments. Facts are the event data confirming the hypothesis.

The 'ladder of inference' show how we jump from some fata to a conclusion only selecting some data and then acting on the conclusion you draw. That action changes future events which get fed back into your reasoning. This is a positive feedback loop, a vicious circle where our beliefs impact how we select data from reality to focus on; selecting only the data which supports our beliefs and directs our behavior**.** One way to avoid mistaken conclusions made on the ladder of inference is test the assumptions which lead us to move up the ladder of inference and the data you are using.

The scientific method moves to formulate hypothesis or explanations of the data and those explanations are tested by making a predictions. Neither the hypothesis nor the prediction describes a real fact. The occurrence of a fact matching the prediction is confirming evidence for the hypothesis. The fact in this scientific approach is neither the hypothesis nor the prediction. Frequently however, in reasoning based on virtual information the hypothesis and/or the prediction is used as if it were a fact, used to generate new conclusions --"pseudo-facts". These "pseudo-facts" are given the same credibility as scientifically tested facts and are used to direct our behavior.

A variation of the ladder of inference problem using unverified data or unexamined relating assumptions is the epistemic equation of causation and correlation. This mistake is a danger in the methodologies advocated in big data science. Before examining the correlation problem we need to look at the place of big data science and analytics in the development of scientific inquiry.

8. ONE TAXOMONY OF SCIENTIFIC REASONING

There have been several characterizations of the methodology of big data science. Their emphasis on correlation is part of what has been called the fourth paradigm of knowledge acquisition. In *The Fourth Paradigm* [12] by Microsoft research, the editors argue that data science involves a fourth paradigm for extending knowledge. This new approach is needed because gathering data is so easy and quick that it exceeds our capacity to validate, analyze, visualize, store, and curate the information. The Fourth Paradigm addresses this challenge—and takes advantage of the opportunities it presents.

They describe the first three paradigms for advancing knowledge (scientific investigation) as:
1. Empirical observation and experimentation, empirical science which describes natural phenomena
2. Analytical and theoretical approaches using models and generalization
3. Simulation or computational science. Within the last 50 to 70 years, the third paradigm of computational science has developed to simulate complex phenomena

The editors' thesis is that although empirical, analytical, and simulation methods have provided answers to many questions, a new scientific methodology driven by data intensive problems is now emerging—the "fourth paradigm."

4. The fourth paradigm, also "known as eScience, has developed to unify theory, experiment, and simulation… data and analysis interoperate with each other, such that information is at your fingertips for everyone, everywhere."

This methodology depends on technology advances in databases, analytics which facilitate the gathering, visualizing and organization of data in such a way to help a limited human mind address the overwhelming quantity of data. Data analytics provide prompts to users for possible interpretations of the data which may not have been noticed by human analysts.

On way to look at this is that in the Fourth Paradigm data analytics prompts fill in the same gaps our belief systems did in the faulty ladder of inference and data analytics is in the same need of support as moving along the ladder of inference.

9. THE CAUSE OF CORRELATIONS

Big data science has been characterized as the triumph of correlations [20]. Correlations, defined as the statistical relationship between two data values. Big data science searches massive data stores looking for correlations between elements that have a high statistical relationship.

In science these correlations are useful as heuristic devices within the sciences. Spotting that fact that when one of the data values changes, the other is likely to change too, is the starting point for many a discovery. However, scientists have typically mistrusted correlations as a source of reliable knowledge in and of themselves, chiefly because they may be spurious – either because they result from serendipity rather than specific mechanisms, or because they are due to external factors [18]. The first thing to note is that although big data is very good at detecting correlations, especially subtle correlations that an analysis of smaller data sets might miss, it never tells us which correlations are meaningful. For example in the USA there is a strong correlation between air conditioner sales and ice cream sales, but no one would argue that buying an air conditioner makes you

want to eat ice cream, or vice versa. Quite obviously, both effects can be attributed to a third factor: hot weather.

Some correlations are humorous in that they obviously do not have a causal explanatory relationship but have a common third cause, e.g. "The Lighting of fasten seatbelt signs in aircraft causes a bumpy ride."[28] Other times the "intelligence" revealing the presence or absence a causal relationship is not that clear and needs to the subject of further 'investigation." Determining causality is extremely difficult in science, and it typically requires experiments that are designed to allow investigators to manipulate the conditions carefully and to rule out any other factors that might be at play.

Predictions should not be based on happenstance. They may work for a while as in Google Flu Trends, but they fail when the context or provenance changes. Without a causal connection one should not generalize from one event to others. We simply has the illusion of precision. Operating under this illusion that correlation implies causality can lead to dangerous as mistakes

10. INDEPENDENCE IS NOT A VIRTUE

Although the first three paradigms developed sequentially they are mutually dependent and it is a mistake to try to separate them; big data can work well as an adjunct to scientific inquiry but rarely succeeds as a wholesale replacement. There are real world and scientific problems what you cannot solve by crunching data alone, no matter how powerful the statistical analysis; you will always need to start with an analysis that relies on an understanding of the world [19]

On way to look at this is that in the Fourth Paradigm data analytics prompts fill in the same gaps our belief systems did in the faulty ladder of inference and data analytics is in the same need of support as moving along the ladder of inference.

Apparent data science success have had limitations as the data changes or the context changes. Google flu detection program (GFT-Google Flu Trends) based on an analysis of Google searches (virtual data) was initially successful in predicting flu and outperformed Center for Disease Control (CDC) data. But then later was outperformed "using already available (typically on a 2-week lag) CDC data". GFT was substantially improved by combing it with data from other techniques [17]. There were at least two problems with GFT. The first was the decontextualization of the data (inadequate provenance) and the second was a data feedback loop, similar to what I have called misinformation propagation. "Collections of big data that rely on web hits often merge data that was collected in different ways and with different purposes — sometimes to ill effect. It can be risky to draw conclusions from data sets of this kind." [19]

All empirical research stands on a foundation of measurement. Is the instrumentation actually capturing the theoretical construct of interest? Is measurement stable and comparable across cases and over time? Are measurement errors systematic? The core challenge is that most big data that have received popular attention are not the output of instruments designed to produce valid and reliable data amenable for scientific analysis

Correlations are useful but without scientific hypotheses questions to answer, premises may be based on digital shadows and pseudo facts are based on ladder of inference errors.

Spotting that fact that when one of the data values changes, the other is likely to change too, is the starting point for many a discovery. But it is a starting point.

We can still learn useful things and make causal inferences from a well specified model testing clearly defined hypotheses using Big Data. We can even run experiments with Big Data: randomizing treatment across respondents, and observing outcomes. [22]

Current software does not do this or encourage this.

11. PROFESSIONAL ETHICS AND BIG DATA SCIENCE

In order to improve the value of big data projects and squeeze more actionable information out of these types of data we need ways to help professionals reduce the occurrence of these "pseudo-facts" and reduce the associated difficulties caused by acting on unreliable or false beliefs

We can apply a precise scientific filter to "messy" data. "If we can have all the data on a specific phenomenon, then surely we can focus on understanding it to a high level of precision, if we so wish? [18] Big data certainly do enable scientist to spot patterns and trends in new ways but the ability to explain why a certain behavior obtains is still very highly valued - arguably over and above the ability to relate two traits to each other. Correlations are but a starting point to a scientific explanation on which we can base predictions.

Basic science techniques could be assigned to critical big data results. There should be new practices established, analogous to double-blind tests that help prevent big data scientists from being misled? There could be multiple groups developing code to analyze big data that remain completely insulated from each other in order to arrive at independent results

We should follow the same sorts of careful attention to the requirements of causal inference that we would follow with any observational data set, we can draw causal inferences. We can still learn useful things and make causal inferences from a well specified model testing clearly defined hypotheses using Big Data. We can even run experiments with Big Data: randomizing treatment across respondents, and observing outcomes. [22]

This approach will help achieve professional goals from the Code of Ethics.

The ACM Code of Ethics from 1992 states that computer professionals should:

1. take "precautions to ensure the accuracy of data,"
2. protect "it from unauthorized access or
3. accidental disclosure to inappropriate individuals."
4. establish procedures "to allow individuals to review their records and correct inaccuracies."
5. define retention and disposal periods for information. [1, Imperative 1.7]

Big data science provides the intelligence which is improved when we add investigation

12. ACKNOWLEDGMENTS

Our thanks to David Longman for extended discussion of the causation correlation relationship and to reviewers of the initial abstract of this paper.

13, REFERENCES

[1] ACM Code of Ethics and Professional Practice http://www.acm.org/about/code-of-ethics

[2] Anderson, C. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine* (June 2008).

[3] Argyris, C. 1990. *Overcoming Organizational Defenses: Facilitating Organizational Learning*,Allyn and Bacon 1st Edition,.

[4] Bollen, J., Mao, H., Zeng, X.J. 2011. Twitter mood predicts the stock market. J. *Comput. Sci*. 2(1), 1–8.

[5] Carr, N. 2015. *The Glass Cage: Where Automation Is Taking Us*, The Bodley Head (Penguin Random House Company) London.

[6] Ciulla, F. et al., 2012, Beating the news using social media: the case study of American Idol. EPJ Data Sci. 1:8 (2012). doi:10.1140/epjds8 http://www.epjdatascience.com/content/1/1/8

[7] Collins, J. 2010. Sailing on an Ocean of 0s and 1s. *Science Books Et Al* 327 (19 March 2010) www.sciencemag.org

[8] Fang, J., Wang, W., Zhao, L., Dougherty, E., Cao, Y., Lu, T., and Ramakrishnan, N., 2014. Misinformation Propagation in the Age of Twitter, *IEEE Computer, 2 (Feb 2014)* pp 90-94.

http://www.computer.org/cms/Computer.org/ComputingNow/issues/2015/02/mco2014120090.pdf

[9] Few, S. 2012. Big Data, Big Ruse, Perceptual Edge. *Visual Business Intelligence Newsletter* (Jul/August/September 2012). http://www.perceptualedge.com/articles/visual_business_intelligence/big_data_big_ruse.pdf

[10] Gantz, J. and Reinsel, D. 2011. Extracting Value from Chaos. *IDC IView, EMC Corporation* (June 2011). www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

[11] Gotterbarn, D. 1999. Privacy lost: The Net, autonomous agents, and 'virtual information'. *Ethics and Information Technology* 1 (pp 147-154).

[12] Hey, T., Tansley, S., and Tolle, K. (Eds.). 2010 *The Fourth Paradigm: Data-Intensive Scientific Discovery*: Microsoft Research, http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[13] Howie, P. 2006. Working with the Ladder of Inference. *ANZPA Journal* (15 December 2006). http://aanzpa.org/system/files/ANZPA_Journal_15_art09.pdf

[14] Intel, 2012 *Big Data 101: Unstructured Data Analytics, (June 2012).* http://www.intel.com/content/dam/www/public/us/en/documents/solution-briefs/big-data-101-brief.pdf

[15] Kastens, K. 2012 Is the Fourth Paradigm Really New? *Earth & Mind: The Blog* (Oct 20 2012) http://serc.carleton.edu/earthandmind/posts/4thpardigm.html.

[16] Lanier, J. 2013. *Who owns the Future?* Simon and Schuster New York.

[17] Lazer, D., Kennedy, R., King, G., Vespignani,A. 2014. Big Data: The Parable of Google Flu: Traps in Big Data Analysis. Science 343 (14 Mar 2014) pp2103-1205 www.sciencemag.org http://scholar.harvard.edu/files/gking/files/0314policyforumff.pdf

[18] Leonelli, S. 2014. What Difference Does Quantity Make? On the Epistemology of Big Data in Biology. Big Data & Society 1 (Jun 2014) ., Sage Publications Inc. DOI: 10.1177/2053951714534395 https://ore.exeter.ac.uk/repository/bitstream/handle/10871/16163/BigData%26Society_Final_April2014.pdf?sequence=2 .

[19] Marcus, G. and Davis, E. 2014. Eight (No Nine!) Problems with Big Data, . (April 9 2014) http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=2

[20] Mayer-Schoenberger, V., and Cuckier, K., 2013. *Big Data: A Revolution That Will Transform How We Live, Work and Think.* John Murray Publisher, London.

[21] La Granga, M. 2007. Megan's Law Listing may be ties to slaying in *LA Times.* Dec 10, 2007. http://articles.latimes.com/2007/dec/10/local/me-molester10

http://claycord.com/2011/10/31/the-megans-law-database-do-a-quick-check-before-trick-or-treating/

Slaying http://on-murders.blogspot.com/2007/12/megans-law-listing-may-have-led-to.html

[22] Nagler, J and Tucker, A. 2015. Drawing Inferences and Testing Theories with Big Data. *PS: Political Science & Politics*. 48, 01 (Jan 2015) pp 84-88. doi:10.1017/S1049096514001796. http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=9492554&fileId=S1049096514001796

[23] Pigliucci, M. 2009. The End of Theory in Science. *European Molecular Biology Organization reports.* 10 (6 June 2009).

[24] Taleb, N. 2010. *The Black Swan: The Impact of the Highly Improbable Fragility (Incerto*) Random House New York.

[25]Tucker, P. 2014. *The Naked Future: What Happens in a World That Anticipates Your Every Move?*. Penguin Group New York.

[26] Tumasjan ,A. et al., 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment in *Proceedings of the 4th International AAAI Conference on Weblogs and Social Medi*a, Atlanta, Georgia, 11 to 15 July 2010 (Association for Advancement of Artificial Intelligence, 2010)

[27] Vickers, G. 1965, *Art of Judgment*, Chapman & Hall, London.

[28] Vigen, T. 2015. Spurious Correlations. Hachette Books, New York

Cloud Computing: the Ultimate Step Towards the Virtual Enterprise?

Norberto Patrignani
Politecnico of Torino, Italy
and Uppsala University, Sweden
Via S.G.Bosco 4, 10015, Ivrea, Italy
+39 348 731 7676
norberto.patrignani@polito.it

Iordanis Kavathatzopoulos
Uppsala University, Sweden
Dept. of Inf. Technology
Box 337 - SE-751 05, Uppsala, Sweden
+46 18 471 6894
iordanis.kavathatzopoulos@it.uu.se

ABSTRACT
This paper proposes a reflection on cloud computing among users, organizations, policy makers, and providers. In particular the focus is on the social and ethical implications for organizations developing a strategy for cloud computing. Also the new roles and responsibilities of the CIOs are analyzed within the complexity of the stakeholders' network around cloud computing. The cloud opportunities but also the issues of concerns are investigated due to their importance for organizations that are more and more shifting towards virtual enterprises.

Categories and Subject Descriptors
K.4.1 [**Public Policy Issues**]: Ethics, Privacy, Regulation.

K.4.3 [**Organizational Impact**]: Employment.

General Terms
Economics, Human Factors, Legal Aspects.

Keywords
Cloud Computing, Computer Ethics, Business Ethics, CIO, Virtual Enterprise.

1. INTRODUCTION
The debate about the push towards centralized organizations due to the introduction of Information and Communication Technologies (ICT) is not new among scholars and researchers [9]. With cloud computing these centralized architectures reach the global scale with an immense impact on the entire society.

The embedded characteristics of cloud computing are [18]: a) network-based, with broadband networks available in most of the countries; b) computing servers as shared-platforms, with resource pooling and multi-tenancy; c) rapid scalability and elasticity; d) measured / metered services (for "billing" purposes); e) on-demand, self-service.

These characteristics, if not mitigated with proper measures, will push organizations and companies towards a complete delegation of their storage and processing functions to cloud providers, with the increasing risk of losing their autonomy.

Conference'10, Month 1–2, 2010, City, State, Country.

In this paper we propose a roadmap that maybe used by small and medium organizations for finding the right balance between the business pressure towards cloud computing (the "heteronomy" attractive option) and the need of maintaining a reasonable "autonomy" in terms of ICT strategies. In particular it is proposed a reflection around the main ethical dilemmas related to cloud computing and their implications for end-users and small and medium organizations. Then we focus on the role of Chief Information Officers (CIOs) inside organizations and, thanks to cloud computing, how the organization itself is rapidly evolving towards a virtual enterprise. This will open a collection of ethical issues for the society, for policy makers, and for the organizations themselves. In particular for enterprises with a strong commitment towards social responsibility: what kind of corporate social responsibility will apply in this virtual world? [19].

2. THE ETHICAL DILEMMAS OF CLOUD COMPUTING
Of course, for many small and medium enterprises (SME), delegating to cloud providers their storage and processing needs is a very attractive move: it will shift, on their yearly financial balance, the information technology expenses from the "capital expenditures" line (CAPEX) to the "operational expenses" line (OPEX).

The CAPEX way for ICT means buying, installing and maintaining ICT infrastructures and computer rooms, including the costs related to ICT personnel, computer professionals and, in many cases, it is needed even a Chief Information Officer, a CIO. The OPEX option is simpler, it means paying the ICT "bill" of the contracts signed with cloud providers without the need of maintaining an internal ICT organization and a data-center. In particular, with cloud services like Software as a Service (SaaS), all data and processing capabilities are delegated to the cloud provider. How to find the proper balance between these two options? How can a CIO take this difficult decision between "heteronomy" and "autonomy"? How will the role of the CIO change? What are the social and ethical implications of this decision?

The organizations will be under economic pressure to go in the direction of cloud computing but is it the right choice? Among the ethical implications related to this choice, three are really fundamental: *humans*, *data protection*, and the *environment*.

Humans: The social risks and impacts of automation on employment were well known since the beginning of the

computer era: the founder of Computer Ethics, Norbert Wiener, pointed out to the scientific community these issues since the beginning, in the 1950's [22]. Today with the immense advances in artificial intelligence, robotics, and cloud computing those worries are even more intense, also due to the deep social and economic crisis in Western societies [16]. If we concentrate on cloud computing consequences, many researches have celebrated the opportunities and take for granted the shift of jobs from one side (internal data centers) to the other side (external data centers, at the cloud providers' sites) and that the savings in ICT expenses will translate automatically in a better economy and more jobs. But these assumptions, even if theoretically sound, are far of being demonstrated as true. Probably it is true that cloud computing will push upwards the professional levels of ICT people, from simple system administrators they will be driven towards less technical jobs like project management, business change-management, suppliers management. The ICT people are the best candidates to cover roles where information technologies cross business processes, they will help companies in streamlining processes and carefully identify whether or not it make sense to use external providers: they will become real "information workers".

Data protection: The data availability, security and privacy issues are the top areas of concern for organizations dealing with cloud computing's strategic decisions, in particular in public authorities and government agencies [4]. In many cases the cloud computing infrastructure will involve the collection and processing of sensitive and personal data (e.g. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction, etc.) that, at the extreme, could even include genetic information and real-time tracking of people. Scenarios like collection of health measurements by means of sensors, that transmit via a body-area-network those data to cloud providers, that collect them into big data repositories for analysis and visualization services, are almost ready. The ethical concerns related to these scenarios need to be addressed not only by justifying the collection of personal sensitive data, or by complying with relevant legislation, cloud providers need to prepare detailed information on the procedures implemented for collecting, storing, preserving, retaining or destroying data. Organizations (potential cloud users) should include these inquiries to the cloud providers during the design of their cloud strategy.

The environment: The climate change will be one of the main challenges for humanity in the next decades. Also ICT is contributing to CO2 emissions for about the same level of airline industry [6], so a careful investigation of possible reduction of ICT related emissions is very important for the grand challenge of climate change. A recent study estimates that the emissions due to ICT, by 2030, will reach 1.25 Gigatonnes of CO2 (GtCO2) distributed in: 28.8% due to datacenters, 47.2% to end-users devices, and 24.0% to networks. Of course ICT applications can also provide a great contribution to CO2 reduction in sectors like mobility, manufacturing, agriculture, buildings, energy, etc. by means of functional optimization and dematerialization; by 2030, this CO2 reduction due to ICT could reach 12.08 GtCO2 [10]. It looks like the balance is positive: ICT creates 1.25 but reduces 12.08 GtCO2, even if, for a more complete investigation, we will need to analyze the entire ICT life-cycle, including mining of materials, device development and manufacturing, and e-waste management.

There is also a kind of "rebound" effect to take into account: the more the cloud will be easy to access, the more it will be accessed, in particular via Wi-Fi networks. Indeed, between data-centers and networks from one side and the end-users devices on the other side, there is the "last-mile"-network, a network that, in most of the cases, is based on wireless technologies. This section of the connection is consuming a growing amount of energy in the entire cloud picture: the energy consumption of this "wireless cloud" will increase by about five times from 2012 to 2015, corresponding to an increase in CO2 from 0.006 GtCO2 in 2012 to 0.030 GtCO2 in 2015 [3].

On another side, cloud providers are the main actors of one of the fastest growing ICT sectors, they grow economically but also their power consumption is growing, and also their responsibility: they could start an enormous shift in the ICT energy ecosystem, starting from their data-centers, asking for more transparency in their electricity supply-chain. This can be the trigger towards a more sustainable electricity generation for ICT that could drive a more general shift toward renewable energies.

3. ISSUES FOR USERS AND SME

If we now move our focus on the cloud users' side we can see more precisely what are the areas of concern related to cloud computing. For users and SME there are many issues that should be addressed before taking a choice, and when designing a strategy about cloud computing. While for a single user these issues can be easily addressed (for example, for data availability, with local data backups) for a SME the picture can become more complex and requires a careful analysis. In the following we will address some of the most important areas of concern:

- *Governance* (with SaaS service model, the entire ICT, Application, Services, Server, Storage, and Network, is delegated to the cloud provider, but also opens a collection of issues like complete unrecoverable data loss, function creep, lock-ins, and possible abuse of power [17]);

- *De-perimeterisation* (for most organizations there will be a loss of "perimeterisation", the traditional boundaries between systems and organizations will disappear; the "virtual" enterprise is becoming a reality);

- *Contractual obligations* (if everybody can buy any amount of ICT resources, then there will be powerful organizations that will buy resources just for re-selling them with a little added-value, they will act like ICT "brokers", then who will be responsible for what? ICT brokers are a very good example of a totally "virtual enterprise");

- *Problem of many-hands* (this means that too many administrators control critical resources; for example, what will happen if a cloud administrator, for maintenance purposes, decides to shut-down a service?);

- *Risk management and reliability* (if something goes wrong – and in ICT this is unavoidable – will they need to trace the events, with a kind of cloud traceability? About reliability: all the issues related to software engineering, from the limits of software reliability, to the responsibility of software designers, are still there, they just moved into the center of the "cloud");

- *Compliance* (cloud users will need to know the data location, for example, for compliance purposes; some users (e.g. financial organizations, public authorities, etc.) need to know which laws apply, in which country are their data);

- *Open market* (if a cloud user organization will want to move to another cloud provider, what kind of freedom to change provider will be available? How will be avoided the risk of monopolies and "lock-ins"? This lock-in is one of the most critical risks associated with the use of cloud computing in enterprise environments and the ongoing consolidation process in the online services industry is confirming this preoccupation).

Addressing all this kind of issues is very important from a decision maker point of view, and may require special attention when organizations define the cloud service contracts.

An interesting approach for addressing the above issues is emerging in the public authorities domain: it is called "community cloud". This solution enables the consolidation of many local small data-centers into one single regional data-center that delivers cloud services to local public administrations. Since this "community cloud" is restricted to authorized agencies only, it avoids the risk of delegating to an external cloud provider the handling of sensitive data about citizens and of data that could be of strategic importance for national security. The European Network and Information Security Agency recommends "community cloud" solutions, for security and resilience reasons, to government agencies [4]. But not all organizations share the same administrative domain like the public agencies. In general, organizations will have to rethink their ICT strategies in a context where cloud solutions are very attractive but bring also many concerns.

As already noted, the CIO, that is the role in the organization with the highest level of knowledge about ICT, will have to change the role: from a simple ICT provider to the responsible for ICT governance. How can decision makers be supported by CIOs in this new complex socio-technical scenario?

4. IS THERE STILL A ROLE FOR CIO?

Cloud computing, as mentioned before, is not only a new technology, it is one of the main paradigm shifts in the history of computing with an immense impact on the processes and on the organizational side of companies. For example, the very simple interface offered by cloud providers is enabling many line-of-business managers inside organizations to buy SaaS software autonomously, bypassing the CIO. This creates a lot of conflicts and expose the company to the risk of a poor ICT management. The role of CIOs in these scenarios it is increasingly important, even if the number of people they manage is decreasing, they still are the main point of coordination of decisions about cloud computing and before taking these important decisions it is recommended to build the most complete related stakeholders' network [13][14].

Probably the most difficult ethical dilemmas will be faced by CIOs. They will have to provide a strategy to companies willing to take the cloud as a serious opportunity. In a way the CIOs will be the cross point between "computer ethics" issues and "business ethics" issues. The CIOs will become the *ethical decision maker* for cloud computing strategies [13].

The old theory that considered the maximization of profits as the sole task of a company, also known as shareholders' theory [8], is no more enough in the complex XXI century's scenario. Many customers and investors are asking for more transparency and the companies need to be managed taking into account the interests of all stakeholders, not only shareholders, but also employees, the community, the environment, the consumers and the society in general. It is the stakeholders' theory [7]. For companies with a strong Corporate Social Responsibility strategy, with a strong commitment to all stakeholders, and willing to adopt transparency towards customers and users, the ethical dilemmas related to cloud computing are fundamental for their future. The shift towards cloud computing is not just an ICT choice. It has immense consequences in terms of organizational level, customer services, and reputation. A strong CSR implies a careful control of the company's "borders", with particular attention to the suppliers.

Here the role of the CIO becomes strategic. Of course there is not a simple formula to find the right answer in this complexity, so we will need to focus on what philosophy is about: the way we think and how to succeed thinking in the right way, i.e. philosophizing. Recalling Plato, leaders (in this case CIOs) are, by definition, not the persons that *have* the right answers, but the persons that are *able to find* the right answers [20]. The CIO will need to become a philosopher!

A useful tool in these cloud computing complex scenarios is the stakeholders' network, where nodes are the stakeholders, and connections are the relationships among stakeholders. For example a company can access a cloud provider's services by means of the network. These services can be subject to norms established by policy makers. The network providers and the cloud providers are using ICT and networks that need to be designed and powered. ICT designers are computer professionals that have their deontology to respect. The power has to be provided by utilities that have a commitment to the environment, environmental advocacy organizations will carefully scrutinize cloud providers asking for about the source of energy powering their data-centers [11], etc.

If we also include in this network the CIO's organization with all its components (the people working in the old data-center, the employees, the customers, etc.) we see the importance of an ethical decision making process for the CIO [14].

5. TOWARDS THE VIRTUAL ENTERPRISE?

If a company can be modeled as a collection of *processes*, *technologies*, *people*, and physical *places*, cloud computing is lowering the importance of all these component but the first one: *technologies* are easily available from the cloud, *people* may be hired on demand via global brokers and the physical *places* where people collaborate are becoming virtual spaces. Only the first component, *processes*, representing the deep identity of the company will survive to this shift towards virtualization.

As mentioned before, cloud computing is attractive in particular for SMEs for going towards *ICT-as-a-Service*. On another front, cloud computing is also an enabling factor for creating a network of SMEs: a virtual enterprise. For example in Europe this opportunity of creating a virtual enterprise as a network-of-collaborating SMEs is a unique opportunity for building a greater business ecosystems, since individually all SMEs have not the critical mass for competing at global scale or, even more difficult, to have the innovation capabilities needed in highly competitive environments. One EU-funded research project, finished in 2014, has addressed exactly this issue: the Business Innovation in Virtual Enterprise Environments (BIVEE). It has supported the creation of an ICT infrastructure for "*... supporting enterprise innovation for networked SMEs and virtual enterprises*" [1].

The light bindings enabled by cloud computing is the perfect enabler for a virtual enterprise where the components (SMEs) made a temporary alliance, based on network connections, for facing a challenging business scenario. They can complement each other competencies, build cross-boundary platforms (see above the *de-perimeterization* due to cloud computing), grow independently from the physical location (data and information can flow over the networks), adopt a *participatory design* approach where a peer-to-peer paradigm is dominant, and, most important, this does not require the creation of a legal entity, it is just a tactical and temporary alliance [2].

In the XXI century's scenario, while the so called "virtual enterprise" is more and more taking shape, many questions are arising, like: What is the identity and reputation of a virtual enterprise? What is the corporate social responsibility of a virtual enterprise?

Here the suggestion is to adopt the *complex systems* paradigm. If we adopt the concept of a complex systems as a collection of a large amount of entities communicating each other, exchanging information, capable of learning from experience and where the single component does not explain the behavior of the whole system (the classic reductionist approach), the only chance we have is to observe a complex system and, in particular, to try to capture its *emerging properties*. For example the emerging property of an elastic object is elasticity even if the single molecule (the component) is not elastic. In a virtual enterprise we have a complex network of organizations, exchanging information, and if we look precisely at one of the components (the single organization) we do not understand the behavior of the entire system. So what is the *emerging property of a virtual enterprise*? Can we accept the concept of *ethics as an emergent property of a complex system* composed by many social agents? [15]. This is an interesting area of research for investigating a loose alliance like a virtual enterprise, where the ethics of this complex (social) system cannot be the collection of ethics of the single components and cannot be a structured set of norms since there is not a legal super-entity, a binding meta-organization's code of ethics.

In the meantime, the virtual enterprise resulting from a collection of *processes*, that, by means of cloud computing *technologies* is connecting *people* in virtual *workplaces* could start by exposing the main components of this network and taking the responsibilities for all the components (considering them as the supply-chain of a real enterprise). This means that customers and investors will ask the virtual enterprise to be compliant at least with the international standards and guidelines like:

- European Policy on Corporate Social Responsibility: "*... To fully meet their social responsibility, enterprises should have in place a process to integrate social, environmental, ethical human rights and consumer concerns into their business operations and core strategy in close collaboration with their stakeholders*" [5];

- UN Global Compact: "*... aiming to create a sustainable and inclusive global economy that delivers lasting benefits to all people, communities and markets*." [21];

- ISO 26000: "*... business and organizations do not operate in a vacuum. Their relationship to the society and environment in which they operate is a critical factor in their ability to continue to operate effectively. It is also increasingly being used as a measure of their overall performance*" [12].

6. REFERENCES

[1] BIVEE 2014. *Business Innovation in Virtual Enterprise Environments*. EU-FP7 Contract Number FP7-ICT-285746. http://bivee.eu/

[2] Camarinha-Matos, L., M., Afsarmanesh, H. 2007. A Comprehensive Modeling Framework for Collaborative Networked Organizations. *Journal of Intelligent Manufacturing*. 18, (Jul. 2007), 529–542.

[3] CEET 2013. *The Power of Wireless Cloud. Center for Energy-Efficient Telecomnunications*, Bell Labs and University of Melbourne, www.ceet.unimelb.edu.au

[4] ENISA, European Union Agency for Network and Information Security 2011, *Security and Resilience in Governmental Clouds*, www.enisa.europa.eu.

[5] EU Commission 2015. Corporate Social Responsibility. ec.europa.eu/enterprise/policies.

[6] Fettweis, G., Zimmermann, E. 2008. ICT energy consumption – trends and challenges. In *Proceesing of the 11th International Symposium on Wireless Personal Multimedia Communications* (WPMC 2008), Lapland, Finland, 8-11 September.

[7] Freeman, E.R. 1984. *Strategic Management: A Stakeholder Approach*, Boston, Pitman Publishing.

[8] Friedman, M. 1970. The Social Responsibility of Business is to Increase its Profits. *The New York Times Magazine*, 13 September 1970.

[9] George, J., F., King, J., L. 1991. Examining the Computing and Centralization Debate, *Communications of the ACM*. 34, 7 (Jul. 1991), 62–72.

[10] GESI, Global e-Sustainability Initiative 2015, *#Smarter2030, ICT Solutions for 21st Century Challenges*, available at: gesi.org (accessed 9 June 2015).

[11] Greenpeace 2012. *How Green Is Your Cloud?* www.greenpeace.org

[12] ISO 2015. International Standard Organization, *ISO 26000 - Social Responsibility*. www.iso.org.

[13] Kavathatzopoulos, I. 2012. Assessing and acquiring ethical leadership competence. In G.P. Prastacos, F. Wang and K.E. Soderquist (ed.), *Leadership through the Classics. Learning Management and Leadership from Ancient East and West Philosophy*. Springer, Heidelberg, 389-400.

[14] Laaksoharju M. 2010. *Let us be philosophers! Computerized support for ethical decision making*. PhD Thesis, Department of Information Technology, Uppsala University, Uppsala.

[15] Minati, G. 1995. Detecting Ethics in Social Systems. In K. Ellis, A. Gregory, B.R. Mears-Young, and G. Ragsdell (ed.), *Critical Issues in Systems Theory and Practice*. Plenum Press. New York, 697-701.

[16] MIT-TR 2015. Who will own the robots? *MIT Technology Review*. July/August 2015.

[17] Mosco, V. 2014. *To the Cloud: Big Data in a Turbulent World*. Paradigm Publishers.

[18] NIST 2011, *The NIST Definition of Cloud Computing*. National Institute of Standards and Technology, NIST Special Publication 800-145, (Sep. 2011), 2.

[19] Patrignani, N., Fakhoury, R., De Marco, M., Cavallari, M. 2015. Cloud Computing: Risques et Opportunites pour la Responsabilite Societale des Entreprises. In *Proceedings of the International Conference on ICT in Organizations*, ICTO2015, Paris, 12-13 March, 2015.

[20] Plato (1992), *Politeia* (The Republic), Athens, Kaktos.

[21] UNGC 2015. *United Nations Global Compact*. www.unglobalcompact.org.

[22] Wiener, N. 1950. *The Human Use of Human Beings: Cybernetics and Society*. 2nd ed. revised, HoughtonMifflin and Doubleday Anchor, Boston, MA, 1954.

Where is Patient in EHR Project?

Anne-Marie Tuikka
Doctoral Student
Information System Science
University of Turku
amstou@utu.fi

Minna M. Rantanen
Graduate Student
University of Turku
minna.m.rantanen@utu.fi

Olli I. Heimo
Project Researcher
Technology Research Center
University of Turku
olli.heimo@utu.fi

Jani Koskinen
Doctoral Student
Information System Science
University of Turku
jasiko@utu.fi

Neeraj Sachdeva
Doctoral Student
Information System Science
University of Turku
neesac@utu.fi

Kai K. Kimppa
Postdoctoral Researcher
Information System Science
University of Turku
kai.kimppa@utu.fi

ABSTRACT
In this paper, we do a literature review on electronic health records (EHR) and patient involvement. It seems that patients are not included as much as one would presume. After our analysis of both literature and ethical nature, we suggest that research on why this is so and whether they should be included needs to be done.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues: Ethics, Regulation, Computer-related health issues.

K.6.4 [**System Management**]: Centralization/decentralization Quality assurance

K.7.m [**The Computing Profession / Miscellaneous**]: Ethics

General Terms
Design, Human Factors

Keywords
Electronic Health Records, Ethics, Development, Implementation, Information Systems, Patient Centeredness

1. INTRODUCTION
The aim of our paper is to set a ground for the discussion about the different aspects of empowering the patient in the electronic health record project (later referred as EHR project). We will approach this topic by studying how patients have been previously involved in EHR projects according to researchers who have studied these projects. Thus, systematic literature review was chosen as appropriate research method for this study. The research question for our systematic literature review is "how patients have participated in design, development, and implementation of their EHR projects according to scientific articles".

Electronic health record (later referred as EHR) is a term which is infrequently and diversely defined and it is often used as synonym for electronic medical record (later referred as EMR). Although equivalence between EHR and EMR has been challenged, terms are still being used as synonyms. (see e.g. [1, 2]) For example International Organization for Standardization (ISO) defines EHR as repository of digital patient data, in which data is securely stored and exchanged and it is accessible by multiple authorised users [3]. They count EMR as one type of EHR, but for example Garets and Davies [1] claim that EHR is a subset of EMR.

In this paper we have chosen to use the term EHR to describe systems that contain personal health information that are used to sustain health of individuals or as part of medical care. Thus, for example personal health records, which are optional for patients but can be used in health care or be directly connected to EHRs [4] are in our scope. EHRs are typically created around patients, but often used by caregivers to stay up-to-date with patient information and progress.

Patients are not only targets or information sources for EHRs, but also users and stakeholders in them. This needs to be kept in mind both when designing and when using EHRs, although, EHRs are often thought as tools for doctors.

This article continues with an ethical discussion about the reasons of taking patients into account during EHR project. Then we present some examples of EHR projects which have involved patients in one way or another. Lastly, we present our preliminary study and discuss its results.

2. PATIENT INVOLVEMENT IN EHR PROJECTS

2.1 Why patient should be involved in EHR Project?
Patients are important stakeholders for EHR projects, because they are both targets and participants of the EHR [2]. As citizens, they have a right to access their electronic health record in many countries. For example, this right exists in the countries of European Union since 2015 [5]. Accessing the information within EHR would give patients a possibility to view their medical

history and to trace the changes in their health [6] which would enable them to do independent health care choices and to participate in the decision-making regarding their medical treatment [7].

From ethical perspective, justification for requirement that patients are considered as core stakeholder of EHR projects can be derived from e.g. the following different ethical premises.

First, the phenomenological and hermeneutical approach to health, healthcare and healthcare information systems requires that healthcare responds to the personal needs of the patients [8, 9]. Those needs are based on individual experiences of patients and thus patients must be heard and be taken into account when developing EHR's which are the core tools of modern healthcare.

Secondly, the justification can be derived from the Four Principles of medical ethics – beneficence, non-maleficence, autonomy and justice – which are the common basis for medical ethics [10]; even though there are critiques towards it, e.g. Lee [11] states that Four principles it is actually based on common moralism or even moral relativism not on moral objectivism like B&C claims [11] and there also exist different ethical codes for healthcare and medicine e.g. WMA Declaration of Helsinki[1]. However, if autonomy and justice are seen as core principles it is an odd situation that the patient – who is the prime target and real source for the existence of healthcare – is bypassed when designing the tools for healthcare. Thus it seems that if autonomy and justice exits they are buried under paternalism and technical determinism.

2.2 How patient can be involved in EHR Project?

This section demonstrates through two examples, how patients can participate in EHR projects. First example is from Finland and it was conducted in collaboration between research institution, public organization and private company. The second example is a project which was funded by European Union.

Work Informatics research group in the University of Turku collaborated with the city of Turku and private software vendor to design a web-based portal for EHR system which were used by the local health centres and citizens of city of Turku. This project followed citizen-centric approach in defining functionalities for the system. In practice, this meant that the project gathered information about the patients need and wishes for using web based portal for EHR system. For gathering this information, cardiac patients were chosen as a target group. Most of them were contacted through invitations, which were posted to them alongside confirmation letter for the reservation to the doctor's appointment to the health centre. 34 cardiac patients wanted to participate to the study and 33 of them could be interviewed at least once during the project. [12.]

The first semi-structured interviews included questions about the participant's background, opinions about current state of digital health services within the city of Turku and wishes for new kinds of digital health services. In addition to interviews, participants were asked to fulfil three surveys about their personal background, quality of life and technological abilities. The interview results were used to develop a new web portal for the EHR systems of the local health centres. This portal was developed by the co-operating software vendor. After the first version of the portal was ready, the participants were invited to test it to the facilities of the software vendor and 17 participants joined the testing. Usually few participants tested it at the same time. Project researchers followed the testing and helped the participant when necessary. After the testing period, each participant was individually interviewed about their experiences of the portal and their ideas for further development. These interviews were semi-structured and they were conducted either on the testing day or few days later. After analysing gathered data, project researchers created a design for a portal which allows citizen to see certain pieces of their health information restored to the EHR system of the local health centres. Such information included reservations to doctor's appointments, past and ongoing medication, and results for laboratory results. [12.]

European Union aimed to facilitate the interconnectedness of national EHRs by launching a project called epSOS. The aim of this project was that EU citizen's health records, such as medical subscriptions in an EU country could be accessed from another EU country. Citizens were invited to join the project for evaluating its outcomes. Every time a citizen had received epSOS service (for example a visit to pharmacy) they were given an evaluation form to be fulfilled. The evaluation form was a paper questionnaire which was instructed to submit to the service provider who forwarded it to epSOS project team for further analysis. [13.]

3. PRELIMINARY STUDY ABOUT THE PATIENT INVOLVEMENT IN EHR PROJECTS

3.1 Research method

In this section, we will present our preliminary study about the patient involvement in EHR projects which is conducted in the form of systematic literature review. The research questions for our systematic literature review are the following:

- How is the patient discussed in academic studies on EHR projects?
- How the patient has been taken into account in EHR projects reported by academic studies?

In order the find suitable articles for our analysis, we aimed to do an algorithm which would return academic articles about EHR projects. Because different authors refer to EHR with different terms [1, 2], we decided that our search would include all the synonyms of EHR identified by our research group. In addition to EHR or its synonyms, our algorithm includes terms which refer to project or certain project phases. We also included terms which refer to patients, because we aimed to find articles which discuss patient related issues in the context of electronic health record project. The final version of our search algorithm is presented below:

("Patient record*" OR "Electronic medical record*" OR "Hospital Information System*" OR "Electronic health record*" OR "Patient information system*") AND (project* OR design* OR develop* OR implement*) AND (customer* OR patient*)

To test our search algorithm and our analyses method, we have restricted our literature search to one database called ABI/Inform Complete and to time span of two and half years between 1st of January 2013 and 22nd of May 2015. Additionally we limited our search to scholarly articles which can be published either in academic journals or scientific conferences. With these limitations, our search algorithm returned 1090 articles.

[1] http://www.wma.net/en/30publications/10policies/b3/

Five authors acted as reviewers and were randomly assigned articles to review. Each article was analysed by two independent authors. Most often, the analyses were made based on the title or the abstract of the article, but the full paper was accessed when the decision could not been made otherwise. The aim of the analysis was to classify the articles to one of the following categories:

1. irrelevant article
2. article relates to health information systems but does not discuss EHR projects
3. article discusses EHR projects
4. article discusses patients' involvement in EHR projects

If the article belonged to the first or the second category, it was not relevant in answering the research questions. If the article belonged to third or fourth category, it was included in the further analysis. Editorial notes, commentaries, book reviews and interviews were excluded from further analysis alongside those articles which could not be accessed in entirety or which were not written in English.

Our initial analysis found 82 articles which discuss EHR projects. Of these, 43 discuss patients' involvement in EHR projects from one viewpoint or another. In the case of 33 articles, the decision was made based on the full paper and 49 papers were assessed based on their abstract. In the case of 10 articles, the decision could not be made because the full paper could not be accessed. In the following sections, we discuss the results of the preliminary analysis of these 82 articles in our data set.

3.2 Results

3.2.1 Interconnectedness between electronic health records and personal health records

According to previous studies [1, 2, 3], academics have not yet agreed on the definition of EHR. However, EHR is often defined as an information system which possesses health information about the patients and which is mainly controlled by the employees of the health care provider (see e.g. [3]). Such a definition distinct EHR from personal health record (later referred as PHR), because PHR is often defined as an information system which restores health information about the patient and which is controlled by the patient (see e.g. [14]).

Our study started by accepting these definitions and we chose to limit our systematic literature review to EHR projects. However, we have found articles which indicate that differences between EHR and PHR may become more indistinct in future. Some of these articles highlight the importance of giving patients more control over their health information which is restored in EHRs. For example, Kellerman and Spencer [15] think that patients should have possibility to view, download, and transmit their health information from EHRs. Ingram ad Arikan [16] call for more open and responsible manner of collecting, sharing, and using personal health data for electronic health records. Thus, citizens could oversee the choices made in relation to their personal health data and they could feel that they control it.

Some other articles concentrate on the growing interconnectedness between electronic health records and personal health records. For example, Baird and Ragdu [4] found that personal health records are more likely to succeed if they are connected with electronic health records. One example of such a personal health record is the information system, which was developed for cancer patients for self-reporting their symptoms [17]. Another example is a standalone information system, which is integrated with an electronic patient record used within Norwegian hospital. With this information system, patients can for example schedule their appointments with medical personal, communicate with medical personnel, record their diet, and review their discharge letters.

In relation to these findings, we have decided to expand our inclusion criteria regarding this study. Alongside those articles, which discuss EHR, we have decided to accept articles, which discuss personal health records if these records are connected with EHR.

3.2.2 How are patients represented in the articles about EHR projects?

Alshameri et al. [6] analyse the development and the current state of EMRs. According to them, EMRs shred out important aspects of the patient because their development did not begin with the essential elements of paper based medical records. They argue that patients and physicians have had least impact in development of EMRs, although, they are the ones interacting and using EMRs the most. Thus, current EMRs do not support enough the interaction between patients and physicians. Alshameri et al. [6] recommend that EMRs should include the complete story which patients have told to the physicians about their health and which physicians have used in their diagnosis. This story is essential for patients if they wish to see their development and possible changes in their health. This story is also used by physicians for making informed diagnosis and prognosis. In addition, institutions need this story for legal purposes.

Boonstra, Versluis and Vos [18] have studied the complexity and typical problems of EHR implementation through systematic literature review. Their systematic review identified only one patient related issue which should be considered during EHR implementation. This issue is patient privacy and confidentiality. Because only this one issue was reported, we assume that Boonstra et al. [18] revealed that EHR implementation are rarely studied from patient's perspective. However, this finding was not mentioned by Boonstra et al. [18] and they do not comment, how patients should be taken into account during EHR implementation.

Pazos, Gorkhali, and DelAguila [19] mention that patients can benefit from implementation of EHR to a hospital. According to them, EHR could reduce the time spend in clinical procedures within hospitals, which would lead to lower fees for patients. Despite of these claims, Pazos, Gorkhali, and DelAguila [19] did not measure the impact of EHR implementation neither to the time spent in clinical procedures nor in the costs of care to patients. Their statistical analysis of EHR implementation in hospitals concentrated on the budget for EHR implementation, the amount of technical staff within hospital, the amount of training offered for medical and administrative staff, and to extent of managerial support for EHR implementation.

Rozenkranz, Eckhardt, Kühne, and Rosenkranz [20] recognise that patients can become actively involved in their healthcare through eHealth solutions, which include integrated EHRs. Their systematic literature review on health information-related applications and services aimed to analyse patients' role in these applications and services. They found that growing number of

applications and services is meant to be used by patients and they assume that future research will concentrate on studying integrated applications which in hold life-long record of patient's health information. However, Rozenkranz et al. [20] do not describe how patients did involve in those development projects which were described within the articles in their data set.

Huang and Chang [21] have studied three hospitals which have offered different kind of eHealth services for their patients. One of these hospitals provided their patients an online access to their medical records. The design of this service was inspired by the customer research conducted among the patients of the hospital in question. Otherwise, Huang and Chang [21] do not describe, how patients have been involved in designing, developing and implementing the eHealth services introduced in their article.

3.2.3 How patients have been involved in EHR project?

We found some articles which described in great detail, how patients have been involved in their information system projects. One of these projects was conducted by Leeds Psychosocial Oncology and Clinical Practice Research Group and it is called eRapid – an information system which allows patient to self-report the symptoms and side effects of their cancer and which is integrated with an electronic patient record. This system was designed, developed and implemented in co-operation with research advisory group which constitutes of 14 patient advocates. These patient advocates were chose to the group because they had or they have had cancer. They have volunteered to the group because they had seen its advertisements or because they have participated to some prior study conducted by the research group. [17.]

Grisot, Hanset and Thorseng [22] describe how patient-centred information system was gradually developed within one Norwegian hospital. In the beginning this information system was only used for communication purposes between patients and the personnel within the hospital in question. Later on this information system was integrated with the EHR within the hospital in question. In addition, its usage spread to other health care units as well.

While this information system was developed, patients' involvement in it grew. In the beginning, the developers only discussed with patients about their experiences. After receiving more funding for the development of this information system, developers organized user workshops for interested patients, who were contacted through patients' association. In the last phase of the project, the developers and the Diabetes Association co-designed new services for patients with diabetes [22].

4. DISCUSSION

4.1 Ethical analysis of patients' involvement in EHR projects

Our preliminary study found only few examples of EHR projects which have involved patients in one way or another in the design, development or implementation phase of the project. Thus it seems that the patients' opinions are not in utmost importance while developing EHR systems – or at least when reporting about them.

EHR should be considered to be a system which promotes the health and wellbeing of the patient. Thus the system should be created with these values in mind because the system should reflect to the ethical values it is meant to promote – and vice versa [23].

It is also a tool used by the medical personnel to do their work – promote the aforementioned health and wellbeing [24]. Thus the design for this particular job should be done efficiently and effectively – by the designer to the professionals – by keeping the requirements of the users in the centre.

The system none the less is not only a tool for the medical professionals to use when treating the patients but also a tool for the patients to maintain and understand their health and wellbeing [24]. As the patients are experts of their own wellbeing [12] all the available information should be easily accessed and understood by them as well and – if the option is available – they should be able to use and insert information to the system with ease.

The main difference here is the amount of patient involvement in the health and wellbeing process. When the patient is only a target (and not a user as in former example) in the system use, we cannot discuss about user-centric design while developing EHR. Therefore we should have a new term – e.g. target-centric design – to discuss about this grey area of information system development.

A way to approach target-centric design is either through an ombudsman [25] or through a patient advocacy organisation. Patient advocacy organisations are problematic because they typically approach the issue from a perspective of patients suffering from a specific illness. This – by definition – limits their point of view to access to EHRs by the patients. The ombudsmen however approach the issue from a perspective of patients who can have a wide range of conditions and thus would be better representatives for target-centric design than patient advocacy organisations' representatives.

None the less where possible the best representatives for the patients' need for their health and wellbeing are the patients. The patients should be made clear how, when and why they are being treated and typically have the last word on whether or not the treatment is given. When relating the EHR as a tool for the treatment the patients are merely ever given the understanding on how their treatment is done in accordance to the system – the tool – used to treat them and their consent is merely asked for the use of their private information. However if this is made more open so that the patients have even the possibility to understand how the system works they may have the possibility to give their **informed consent** on the matter.

As shown previously, given the small amount of scientific research we found on the EHRs developed with the patients it is still unclear if a large enough representative group can be selected and how it should be selected when the design process request help in the design and use of the EHR, but according to the former, it seems unethical not to do that research.

4.2 Limitations and future research

In this article, we have presented our preliminary study about patients' involvement in EHR projects. Although, our study has significant limitations, such as narrow amount of literature sources and the time span of two and half years, it has served as an important test case for our research group.

While testing the search terms for our systematic literature review, we noticed that there is a huge amount of literature which relates to EHR in some way or another even if we limited our search to

ABI/Inform Complete. Thus, we decided to use the additional limitation offered by the search engine of the database which enabled us to restrict our search only to scholarly articles. We also decided to abandon all abbreviations from our search terms because they had multiple interpretations. Even with these limitations our original search returned more than 8000 articles. Thus, we decided to use shorter time span which would return us approximately 1000 articles, because this was considered as feasible amount of articles to review by our research group within the given time span of six months.

After reviewing all the articles in our data set, we have come up with essential ideas for elaborating our systematic literature review for future research purposes. Firstly, we need to expand our search to larger time span of five to ten years and to multiple databases, such as MEDLINE, Academic Search Premier, Business Source Premier, Health Technology Assessment Database, and IEEE Explore. To accomplish this, we need to find a way to do our searchers more precise. In our current data set more than 70 % of articles are totally irrelevant for our search which means that they neither discuss the use nor the development of EHRs (or related information systems). Most often these articles are from the field of medicine. These articles were found with our search algorithm, either because EHR might have been used as a tool for conducting the study in question or because EHR was mentioned in the list of references. In future, we plan to restrict our search to title, to abstract and to keywords in order to reduce the amount of irrelevant articles returned by the search. On the other hand, the problem of such limitations is that those articles, which do not use search terms in title, in abstract or in keywords, cannot be found with our search algorithm. Thus, we aim to reanalyse the titles, the abstracts and the keywords of the 75 relevant articles within our dataset in order to find those terms which should be included in our search algorithm.

5. CONCLUSION

In this article, we have described our preliminary study about the patients' participation in EHR projects. In the beginning of our article, we ethically analysed different reasons for which patients should be taken into account during EHR projects. Then we introduced examples of the different ways in which academics have previously discussed patients' role in EHR projects. We noticed that some academic articles recognise patients as stakeholder group of EHR projects where as other do not. There are also some articles in which patients are recognized as possible beneficiaries of EHR implementation, although the impacts of EHR implementation toward patients are not studied.

We also presented few examples of involving patients in EHR projects. These examples served as a starting point for our ethical analysis of patients' involvement in EHR projects. The conclusion of our analysis was, that best representatives for the patients' needs are the patients themselves. The patients should have possibility to understand how their treatment is done in accordance with EHR, thus, they would be able to give their informed consent for using their private information as part of the EHR.

6. REFERENCES

[1] Garets, D. & Davies, M. (2005), Electronic Patient Records EMRs and EHRs Concepts as different as apples and oranges at least deserve separate names. Healthcare Informatics, October 2005.

[2] Rantanen, M., & Heimo, O. (2014). Problem in Patient Information System Acquirement in Finland: Translation and Terminology. ICT and Society: 11th IFIP TC 9 International Conference on Human Choice and Computers, Turku, Finland, July 30 – August 1, 2014.

[3] Häyrinen, K., Saranto, K. &Nykänen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. International Journal of Medical Informatics 77 (2008) 291–304.

[4] Baird, A., & Raghu, T. (2015). Associating consumer perceived value with business models for digital services. European Journal of Information Systems, 24, 4-22.

[5] Official Journal of the European Union, L 88, 4 April 2011.

[6] Alshameri, F., Hockenberry, D., & Doll, R. (2014). The map is not the territory: the missing patient in the electronic medical record. VINE: The journal of information and knowledge management systems, 44(4), 548-557.

[7] Sunyaev, A. (2014). Consumer Facing Health Care Systems. e-Service Journal, 9(2), 1-23.

[8] Koskinen, J. (2010). Phenomenological view of health and patient empowerment with Personal Health Record. Navigating the Fragmented Innovation Landscape: Proceedings of the Third International Conference on Well-being in the Information Society.

[9] Svenaeus, F. (2001). The Hermeneutics of Medicine and the Phenomenology of Health: Steps towards a Philosophy of Medical Practice" (second revised edition): Kluwer.

[10] Beauchamp, T. L., & Childress, J. F. (2013). Principles of biomedical ethics. New York :: Oxford University Press.

[11] Lee, M. J. H. (2010). The problem of 'thick in status, thin in content' in Beauchamp and Childress' principlism. Journal of Medical Ethics, 36(9), 525-528.

[12] Knaapi-Junnila, S., Korpela, A., Koskinen, J. Lahtiranta, J., Otim, R., & Tuomisto, A. (2015). Kansalaislähtöisyys sähköisissä terveyspalveluissa: Sydänpotilaan arki. Turku Centre for Computer Science. TUCS National Publication, No 20, June 2015.

[13] epSOS, European Patients Smart Open Services, Evaluation for Patients. http://www.epsos.eu/evaluation-for-patients.html, accessed 7.7.2015.

[14] Lee M., Delaney, C. & Moorhead S. (2007). Building a personal health record from a nursing perspective. International Journal of Medical Informatics. Volume 76, Supplement 2, October 2007, S308–S316.

[15] Kellermann, A., Jones, S. (2013). What It Will Take To Achieve The As-Yet-Unfulfilled Promises Of Health Information Technology. Health Affairs, 32(1), 63-68.

[16] Ingram, D., & Arikan, S. (2013). The Evolving Role of Open Source Software in Medicine and Health Services. Technology Innovation Management Review, January 2013, 32-39.

[17] Absolom, K., Holch, P., Woroncow, B., Wright, E. P., & Velikova, G. (2015). Beyond lip service and box ticking: how effective patient engagement is integral to the development and delivery of patient-reported outcomes. Quality of Life Research, 24, 1077–1085.

[18] Boonstra, A., Versluis, A., Vos, J. (2014). Implementing electronic health records in hospitals: a systematic literature review. BMC Health Services Research, 14(370).

[19] Pazos, P., Gorkhali, A., DelAguila, R. (2013). The Role of IS Infrastructure and Organizational Context on Implementation of Electronic Health Records Systems. Proceedings of the 2013 Industrial and Systems Engineering Research Conference, 676-681.

[20] Rozenkranz, N., Eckhardt, A., Kühne, M., & Rosenkranz, C. (2013). Health Information on the Internet: State of the Art and Analysis. Business & Information Systems Engineering, 4/2013, 259-274.

[21] Huang, E., & Chang, C. (2014). Case Studies of Implementation of Interactive E-Health Tools on Hospital Web Sites. e-Service Journal, 9(2), 46-62.

[22] Grisot, M., Hanseth, O., Thorseng, A. (2015). Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. Journal of the Association for Information Systems, 15(4), 197-219.

[23] Heimo, O., Kimppa, K., & Nurminen, M. (2014). Ethics and the Inseparability Postulate. ETHICOMP 2014, Pierre & Marie Curie University, Paris, France, 25th – 27th July 2014.

[24] Koskinen J. (2014). "Life of mine" – Datenherrschaft as Heideggerian way to treat individual's patient information. CEPE 2014 Conference- Computer Ethics: Philosophical Enquiry.

[25] Lahtiranta, J. (2014). New and Emerging Challenges of the ICT-Mediated Health and Well-Being Services. Dissertation, Juvenes Print Oy, Turku, 2014.

Operationalising Design Fiction for Ethical Computing

Joseph Lindley
HighWire Centre for Doctoral Training
Lancaster University
United Kingdom
joseph.lindley@gmail.com

Dhruv Sharma
HighWire Centre for Doctoral Training
Lancaster University
United Kingdom
d.sharma2@lancaster.ac.uk

ABSTRACT
Design fiction is a type of speculative design, where story worlds are crafted to then be used as a canvas upon which so-called diegetic prototypes can be sketched [10]. Because these prototypes exist only within story worlds they are not constrained by currently available technology; because of this design fictions are excellent means to open up space for critical conversations about the future [3, 8]. This project experiments with using design fiction as a novel way to explore the complexities of technology and ethics. We focus on one specific case to demonstrate the method we adopted, however the contribution is general in nature and may be applicable to other cases too. The work consists of two parts, this paper and a 'design fiction documentary' film, 'Care for a Robot' [6]. The paper and film are intended to be viewed together.

Categories and Subject Descriptors
K.4.1 [**Computing and Society**]: Public Policy Issues – *ethics*.

General Terms
Design, Experimentation, Human Factors

Keywords
Design fiction, domestic robots, care for the elderly, radical digital interventions, accessible ethics.

1. INTRODUCTION
We are Dhruv Sharma and Joseph Lindley, we are both doctoral students at the HighWire Centre (Lancaster University). Because of the novel format of this work we have included this section to provide some context and make clear what our personal interests are, why we are doing this research, and how we think it relates to ETHICOMP. Joseph is researching the relatively immature concept of design fiction, he's interested in understanding what design fiction's kernel is and the range of ways it can be used. Dhruv is researching loneliness among the elderly. In particular his research is interested in how 'radical and digital interventions' [18] may be used to reduce the negative impacts of loneliness among the elderly.

The example case that this work revolves around is domestic care robots. Although not commercially available at present, current discourse around medical robots designed to care for people leads us to believe that having an accessible and meaningful debate about the ethical implications of these, potentially pervasive, technologies is essential given the breadth of their impact when (or if) they do become viable [15]. The ideas presented here signify our early response to this challenge. How can we prepare for potentially pervasive technologies in the offing? The work isn't intended to be a manifesto or statement of truth about what ethical stance 'should' be adopted vis-à-vis domestic care robots, nor is it intending to posit the 'best' method to address the challenge of preparing for an ethical debate around caring robots. Rather it describes the concepts, theory and practicalities behind one possible way of accessing the debate and making it more meaningful. As such we think this approach may be replicable for other cases, and we also see this as a general contribution to studies of computing and ethics.

1.1 Radical Digital Interventions
Sharma et al.'s review of existing age-related loneliness interventions, highlights that the majority demonstrate an incremental approach to addressing the problem [18]. They argue that 1) there are relatively fewer interventions that are 'radically' different and that 2) use of digital technology is underrepresented in this area. In order to explore possible strengths - or limitations - of this type of intervention we should pay extra attention to radical-digital interventions and strive toward experimentation and innovation in this area.

The distinction between incremental and radical interventions is akin to *reformist versus radical* departures in environmental discourses [7]. Reformist departures seek solutions within familiar modes of rational management, whereas radical departures argue for a comparatively significant movement away from industrial modes of living and being. Manzini suggests that incremental innovations represent our existing ways of 'thinking and doing' whereas innovations falling outside our current ways of 'thinking and doing' represent radical innovation [14]. Norman and Verganti define *incremental* innovation as "improvements within a given frame of solutions" and "doing better, what we already do" but describe *radical* innovation as "a change of frame" or "doing what we did not do before" [16].

Improvements upon 'what we already do' are usually backed up by reflective practices and learning from past experiences. Radical 'changes of frame' however are either the product *of*, or ultimately lead *to,* uncharted territories. It is therefore impossible to predict the ramifications, implications and impact of radical innovation unless we speculate about what forms those innovations may take. Practices such as design fiction offer us with an academically grounded approach to crafting, interpreting, and making sense of these speculations [cf. 2, 9, 12].

1.2 Design Fiction
There are scant arguments for clearly bounding precisely what design fiction is and how it should be used. It has demonstrably

been used as a prototyping tool, research method, ideation aid, and as a communication tool [4, 11, 13, 22]. Design Fictions harness the power of speculative design thinking to holistically imagine how ideas from the present would manifest in the future. Designers and practitioners create design fiction artefacts in a huge range of shapes, sizes and media: film, text, objects, and combinations of all of them [11]. The most popular definition of design fiction refers to the purposeful application of diegetic prototypes to encourage a suspension of disbelief about change [5] (refer to [11, 12] for a more in-depth deconstruction of this definition).

Lindley's pragmatics framework for design fiction proposes some categories of design fiction intended to make communications about applications of design fiction clearer. As part of that work Lindley introduces a nomenclature to differentiate between design fictions that are created from scratch (intentional design fictions) and other entities that coincidentally share the properties of a design fiction (incidental design fictions) [11].

'Anticipatory ethnography' proposes using observations of design fictions as part of design ethnography research projects. Where design ethnographers tend to do 'quick and dirty' ethnographic studies of people and places in order to design things better, anticipatory ethnographers might do similarly quick and dirty ethnographic studies, but of the people and places in a design fiction world as opposed to the real world. A straightforward example of how one might use anticipatory ethnography is to take a piece of science fiction cinema that meets the criteria of being an incidental design fiction. Watch the film to take detailed ethnographic notes of the action and situations, and then to apply methods of design ethnography in order to generate actionable insights pertaining to the world and diegetic prototypes depicted in the film. If the film's ability to suspend disbelief with diegetic prototypes is strong, then anticipatory ethnography should generate powerful insights [12].

1.3 Robot and Frank

We cast the 2012 film Robot and Frank [17] as a piece of incidental design fiction. Set in an unspecified near future where today's modern hybrid cars are aging and rusty, and the local library is finally withdrawing paper books. The film depicts an elderly man called Frank, his children, and the introduction of a caring robot into Frank's life. Some aspects of how the robot interacts with humans in the film might appear unrealistic, however we argue that on the whole the diegetic prototypes in the film *are* able to suspend disbelief about change, and therefore it passes the test of being an incidental design fiction. A full discussion of what can or cannot be considered incidental design fiction is unfortunately beyond the scope of this paper.

Used as a stimulus, Robot and Frank was essential to producing Care for a Robot, however it is not necessary to actually watch the film in order to make sense of the work and take some value from it. However, we personally recommend it as being a simple, yet thought-provoking film, and also to further contextualize this work. Some sequences from Robot and Frank appear in our film Care for a Robot.[1]

1.4 Care for a Robot

This work is slightly unusual in that it has a two-dimensional relationship with design fiction. It *extends* the incidental design fiction that is Robot and Frank, in order to then *create* an intentional design fiction, Care for a Robot. Furthermore the format of Care for a Robot is, as far as we are aware, the first of its kind: a design fiction documentary.

The film was made by first showing selected clips from Robot and Frank that depict interactions between humans and the robot to the contributors who would eventually be the interviewees in Care for a Robot. The clips were selected to be deliberately thought provoking and encourage debate around whether the interactions shown were possible, plausible, or desirable.

The clips were shown to the interviewees, then we briefly introduced the relevant concepts in an informal discussion (including radical digital interventions, design fiction, anticipatory ethnography and our vision for Care for a Robot). Before filming any interviews we then asked interviewees to imagine they were living in a world where caring robots, just like the one they had seen 'diegetically situated' in Robot and Frank, were a reality and that either they or somebody close to them had experience of working with or owning these robots. Through dialogue between ourselves, and the interviewees, we developed a range of scenarios and personas that you see in the finished film. These are varied and include: a prospective customer buying for her father in law; a hacker who wants revenge after her robot's data was commandeered by the manufacturer; an employer who has appropriated care robots in order to access cheap labour; an academic who bought, and then returned, a care robot for his elderly parent.

None of these 'workshop' sessions were longer than 30 minutes. We did not script any of the responses, and the footage you see in the finished film is constructed from entirely improvised or 'off the cuff' responses to interview questions. Care for a Robot is not chronological and instead focuses on highlighting themes that emerged in the interviews.

2. RELEVANCE TO ETHICOMP

The primary purpose for this paper is to present a method for exploring the ethical considerations of radical digital interventions. In our example case the radical digital intervention is a domestic care-giving robot, however we suggest that the same method may be applied to other cases too. Although the method itself is the significant contribution here, we have included some examples quotes from Care for a Robot in the paper too (see 2.3). It is important to stress that we intend this work to initiate a discussion about how to use design fiction as a means to explore ethics as opposed to adopting a didactic position. This work is a first step.

2.1 We Are Not Ethicists

Although it should be clear by this point already, we want to reiterate that we are not ethicists. Neither were the interviewees that feature in Care for a Robot. However we believe that this fact - that could be seen as a shortcoming for a paper about ethics – is not detrimental to the kernel of this work.

Design fictions tend to present the future as mundane. The future is an accretive space that may well include the buzzing of a cathode ray tube screen right alongside the sheen of a super-thin curved 3D-capable display. In Robot and Frank, rusty and ageing first generation hybrid cars are depicted sharing the roads with super-modern all-electric models. The future will not be a white-walled utopia but will be inhabited by a menagerie of semi-broken technologies and protagonists that, as we do today, are mainly motivated by everyday considerations [1, 24]. By leveraging the

[1] The copyrighted materials from Robot and Frank are included under 'fair use' as part of a research project.

future mundane (as it's shown in Robot and Frank), filtering those situations through the everyday perspectives of our interviewees, then finally packaging the outcome into a digestible format, is how this work creates meaning and generates value.

Being able to produce and contain insights pertaining to radically different ideas, while encoding the essence of everyday mundanity is how this work proposes to bring something new to the ethicist's toolbox. Because we're trying to tease out the 'warts and all' character of the future scenario being explored, it doesn't appear to be the case that our position as 'non-ethicists' has been too much of a hindrance.

2.2 Ethics and The Future

The challenges of understanding the ethics of technology appear to be necessarily bound to the future. We agree that as regards the ethics of technology "At bottom, these issues reduce to traditional ethical concerns having to do with dignity, respect, fairness, obligations to assist others in need, and so forth" [23]. The core ethical issues tend to remain quite static, meanwhile radical technological advances change the situations that these issues apply to considerably. It is the nature of these innovations, and the specifics of the situations they create, that are the largest challenge for ethicists. Design fictions naturally tend towards developing plausible concepts aligned with the trajectory of change, while also *communicating* these concepts with a high degree of 'situativity' [cf. 12, 21].

Second, if we want to explore these possible scenarios - which of course are plural - then we need a means to ask meaningful 'what if' questions, as well as a means to understand the answers. There are various ways in which one might approach asking these questions [19, cf. 20]. We feel this design-fiction orientated approach has some distinguishing factors. First it has the ability to interrogate technologies radically different to those currently available; second that it does so within the brackets of a future mundane; third the ideas contained in the design fiction stimulus are filtered by the everyday and human responses of the interviewees. This results in insights that we refer to as 'diegetically situated'.

2.3 Example Quotes

We are clear that this work's primary aim is to describe and advocate for using design fiction as a tool to open a discursive space from which insights about ethics may emerge. As self professed non-ethicists we're tentative about making any direct claims to do with ethical insights. More important than our own interpretations we hope that presenting this work at ETHICOMP 2015 will stimulate discussion and encourage interrogation of the idea such that it may be developed further, perhaps adapted, and hopefully adopted in other projects.

Despite intending for this work to, first and foremost, be a 'jumping off point' for further discussion, we have included a small selection of quotes from the interviews in the film in order to highlight some provocative examples.

2.3.1 Price vs. Value

Quite separately from the monetary value of the robot, or the cost to the user, the interviewees demonstrated a range of differing opinions about how to quantify the value of the robot carers.

> *"I would argue that this is a trade-off... it depends on what we would trade off for the services we have"*

This interviewee accepts that the companies providing the robots may take something back in order to offset the cost of the robot, perhaps by monetising the data gathered by the robots. This seems consonant with 'free' services available on the web today, for example Google's suite of applications, or the services made available by numerous social networks.

> *"We have three wonderful kids but they give our sitters a hard time… I know they're not intended to take care of children"*

The interviewee's children are apparently notoriously difficult for baby-sitters to handle, whereas using a robot carer to perform baby-sitting duties – which may be more expensive monetarily – appears to be preferable for her.

> *"We got it as a robot carer and what it was turning into was a research tool for the company"*

During a year-long contract this interviewee became aware that, in accordance with the terms and conditions set out by the service provider, data gathered by the robot would be used in a number of unexpected ways, which are perhaps undesirable, and were not clear at the outset.

2.3.2 How Robot Carers Are Perceived and Used

As well as the intended application – to be domestic care robots for elderly users – some of our interviewees appropriated their robots to do jobs and tasks that way were not, perhaps, intended.

> *"I've found them to be extremely useful as flexible labour"*

An entrepreneur, this interviewee has purchased many robots to work across his service-industry business as a cost-saving measure: human labour is unable to compete in terms of bottom-line hourly cost.

> *"..on the off chance.. if the robot happened to capture information from his medical records.."*

This interviewee remotely reviews logs of the robot caring for his grandfather in order to discern what medication his grandfather is taking. It is unclear whether monitoring this level of detail is done with consent, and whether that was the intended use of this function.

> *"The robots outlook is that 'the best way to take care of elderly people is to have robot carers in their homes'"*

This interviewee has become convinced that the robot caring for his wife's parents is trying to influence their behaviour, by, for instance, arranging their walk times so that they will encounter other people with caring robots.

2.3.3 Service Provision

All of our interviewees assumed that large corporations were providing the robot carers, either in a traditional ownership model or 'as a service'.

> *"We helped them buy a microwave, so they weren't about to go and buy a robot on their own"*

Installing a care robot to care for an elderly relative may-well necessitate dealing with highly technical issues, where the end-user might not be technologically savvy enough to have a full comprehension.

> *"They offer a personalised service... obviously you can't just unbox them and let it go… Somebody goes into his house and monitors his interactions with people so they can pre-program the robot"*

This interviewee is very positive about the pre-sales support and level of personalisation that the company offered to support the installation of a care robot at her father in law's house.

> *"Any 3rd party service providers had to sign a disclaimer [if the robot was in the house]… it's like those messages saying 'this call may be monitored for training purposes'"*

This interviewee was not initially aware that the contract with the robot provider insisted that *anyone* entering the house was required to sign a disclaimer allowing the company to use data gathered during their visit.

3. IMPLICATIONS

First and foremost we would like this work to stimulate a conversation with the ETHICOMP community. Does this design fiction centric approach to opening a discursive space about the ethical implications of radical interventions hold any merit? If so what frameworks could be applied to critically examine design fictions like Care for a Robot?

This work, that considers a Hollywood film as a piece of incidental design fiction, adapts the ideas within anticipatory ethnography, in order to then produce a new design fiction documentary, is a first. By focussing on domestic care robots, and in particular trying to discern insights about the ethical implications of this technology, our approach attempts to bound the design fiction, encouraging the discursive space to converge on around a single theme.

Although we have focussed this work on a single type of radical digital intervention we are keen to experiment with applying this approach to other types of radical innovation, perhaps those that have not been conceived yet.

Finally we would like to understand if applications of design fiction might be complimentary to more traditional research into the ethics of computing. Can the relationship between these disciplines be mutually beneficial?

4. ACKNOWLEDGMENTS

Many thanks to everyone in the closing credits of the film, your help and contributions were invaluable to doing this work. We're grateful to the stars and makers of Robot and Frank, for their film inspired this work. We would like to thank Robert Potts for his help developing our original formulation of 'anticipatory ethnography'. Thank you to our supervisors at Lancaster University. This work was funded by the UK Digital Economy Programme (Grant Reference EP/G037582/1).

5. REFERENCES

[1] A Design Fiction Evening with the Near Future Laboratory: 2013. *http://vimeo.com/84826827*.

[2] Auger, J. 2013. Speculative design: crafting the speculation. *Digital Creativity*. 24, 1 (Mar. 2013), 11–35.

[3] Bleecker, J. 2009. Design Fiction: A short essay on design, science, fact and fiction. *Near Future Laboratory*. (2009).

[4] Blythe, M. and Buie, E. 2014. Chatbots of the Gods: Imaginary Abstracts for Techno-Spirituality Research. *Proc. NordiCHI 2014*. (2014), 227–236.

[5] Bruce Sterling Explains the Intriguing New Concept of Design Fiction (Interview by Torie Bosch): 2012. *http://www.slate.com/blogs/future_tense/2012/03/02/bruce_sterling_on_design_fictions_.html*. Accessed: 2014-02-09.

[6] Care for a Robot: 2015. *https://www.youtube.com/watch?v=VKKlnpNueaY*.

[7] Dryzek, J.S. 2005. *The politics of the earth: environmental discourses*. Oxford University Press.

[8] Dunne, A. and Raby, F. 2013. *Speculative Everything*. The MIT Press.

[9] Hales, D. 2013. Design fictions an introduction and provisional taxonomy. *Digital Creativity*. 24, 1 (Mar. 2013), 1–10.

[10] Kirby, D. 2010. The Future is Now: Diegetic Prototypes and the Role of Popular Films in Generating Real-world Technological Development. *Social Studies of Science*. 40, 1 (Sep. 2010), 41–70.

[11] Lindley, J. 2015. A pragmatics framework for design fiction. *Proceedings of the European Academy of Design Conference* (2015).

[12] Lindley, J. et al. 2014. Anticipatory Ethnography: Design fiction as an input to design ethnography. *Ethnographic Praxis in Industry Conference* (2014).

[13] Lindley, J. and Potts, R. 2014. A Machine. Learning: An example of HCI Prototyping With Design Fiction. *Proceedings of the 8th Nordic Conference on Human Computer Interaction*. (2014).

[14] Manzini, E. 2014. Makings Things Happen: Social Innovation and Design. *Design Issues*. 30, 1 (2014), 57–66.

[15] Medical robotics: the solution for our demographic challenge of an aging population? *http://innorobo.com/medical-robotics-the-solution-for-our-demographic-challenge-of-an-aging-population/*. Accessed: 2015-07-07.

[16] Norman, D.A. and Verganti, R. 2014. Incremental and Radical Innovation: Design Research vs. Technology and Meaning Change. 30, 1 (2014).

[17] Schreier, J. 2013. *Robot and Frank*.

[18] Sharma, D. et al. 2015. Radicalising the designer: Combating age-related loneliness through radical-digital interventions. *Cumulus Conference: The Virtuous Circle* (Milan, 2015).

[19] Stahl, B.C. 2011. IT for a better future how to integrate ethics, politics and innovation. (2011).

[20] Stahl, B.C. et al. 2013. The empathic care robot: A prototype of responsible research and innovation.

Technological Forecasting and Social Change. 84, (2013), 74–85.

[21] Suchman, L. 1987. *Plans and situated actions: the problem of human-machine communication*. Cambridge University Press.

[22] Tanenbaum, J. 2014. Design fictional interactions. *Interactions*. 21, 5 (Sep. 2014), 22–23.

[23] Tavani, H.T. 2011. *Ethics and Technology*. John Wiley & Sons.

[24] The Future Mundane: 2013. *http://hellofosta.com/2013/10/07/the-future-mundane/*. Accessed: 2014-10-07.

Juries: Acting Out Digital Dilemmas to Promote Digital Reflections

Elvira Perez Vallejos[1], Ansgar Koene[1], Chris James Carter[1], Ramona Statache[1], Tom Rodden[1], Derek McAuley[1], Monica Cano[1],Svenja Adolphs[2], Claire O'Malley[3],
[1]Horizon Digital Economy Research Institute, University of Nottingham, [2]School of English, University of Nottingham, [3]Faculty of Science, University of Nottingham Malaysia Campus
{elvira.perez, ansgar.koene, christopher.carter, ramona.statache, tom.rodden, derek.mcauley, monica.cano, svenja.adolphs, claire.omalley}@ nottingham.ac.uk

Kruakae Pothong[4], Stephen Coleman[4]
[4]School of Media and Communication. Faculty of Performance, Visual Arts and Communications, University of Leeds, UK
{cskp, s.coleman}@leeds.ac.uk

ABSTRACT
A quick journey through prevention science (e.g., substance misuse prevention) and a comparison between online and offline risks, harm, and vulnerability in children suggests that new approaches and interventions are needed to promote Internet safety and minimise the new sources of risk associated with accessing the Internet. In this paper we present a new methodological approach to promote digital literacy and positively influence the way in which young people interact with the Internet: *iRights Youth Juries*. These juries offer a solution for the challenge of how to engage children and young people in activities that, rather than simply promoting Internet safety, aim to provide the knowledge and the confidence required for developing healthy digital citizens. This approach thus begins to move beyond the notion of the Internet as a simple cause of social change, approaching it instead as an opportunity to engage knowledgeably with the digital world and maximise citizenship.

Categories and Subject Descriptors
J.4 [**Social and Behavioural Sciences**]
K.4 [**Computers and Society**]: Social Issues

General Terms
Measurement

Keywords
Digital rights, Internet safety, children and young people, vignettes, drama, education, engagement.

ETHICOMP 2015, September 7–9, 2015, De Montfort University, Leicester, UK.

1. INTRODUCTION
The Internet is frequently held to transform social relationships, the economy, vast areas of public and private life across all ages and, probably very soon, across all cultures. Such arguments are often recycled in popular debates, sensational tabloid news materials, and indeed in academic contexts as well. Research discussions on the topic of the Internet oscillate between celebration and fear, where on the one hand, technology is seen to create new forms of community and civic life, and to offer immense resources for personal liberation and participation, while on the other, it poses dangers to privacy, creates new forms of inequality and commercial exploitation, in addition to increasing individual exposure to addiction triggers, abuse, and other forms of harm.

These kinds of ideas about the impact of technology tend to take on an even greater force when they are combined with ideas of childhood and youth. The debate about the impact of media and technology on children has always served as a focus for much broader hopes and fears about social change. On the one hand, there is a powerful discourse about the ways in which digital technology is threatening or even destroying childhood. Young people are seen to be at risk, not only from more obvious dangers such as pornography and online paedophiles, but also from a wide range of negative physical and psychological consequences that derive from their engagement with technology. Like television, digital media are seen to be responsible for a whole range of social ills—addiction, antisocial behaviour, eating disorders, educational underperformance, commercial exploitation, depression, envy and so on.

In recent years, however, the debate has come to be dominated by a very different argument. Unlike those who express regret about the media's destruction of childhood innocence, advocates of the new "digital generation" regard technology as a force of liberation for young people—a means for them to reach past the constraining influence of previous generations, and to create new, autonomous forms of communication and community. Far from corrupting the young, technology is seen

to be creating a generation that is more open, more democratic, more creative, and more innovative than that of their parents.

Taking into account both the risks and opportunities associated with the Internet and digital technologies, this paper considers the unavoidable dialectical in which the Internet is both socially shaped and socially shaping. In other words, by studying the way in which the Internet is utilised we gain insights into its overall role and impact, but we also uncover its inherent constraints and limitations which are in turn largely shaped by the social and economic interests of those who control its production, circulation, and distribution. Understanding the values and ideas that are encoded in and promoted through the structure and use of the Internet is essential for successfully managing the social, economic, and cultural effects that it generates.

2. INTERNET SAFETY

At present, there appears to be little robust research evidence that compares the success of available Internet Safety programs, or examines what materials or educational approaches are cost-effective, and how programmes are being implemented in the community. Outcome evaluations have been limited in sophistication, and so far current results show little evidence that Internet Safety programmes reduce risky online behaviours or prevent negative experiences. On the contrary, studies have indicated that while children within test groups are able to retain the extra knowledge presented to them, the learning has been found to have little impact on children's online behaviour [1].

In response to increasing concerns about the extent to which Internet activities put children and young people at risk from sexual and psychological abuse, numerous Internet safety educational materials including online guidelines, tools, and advice for parents and teachers have been developed with the intention of minimising such risks. Internet Safety, however, appears to have more in common with risk prevention programmes than programmes aiming to promote digital rights among children and young people. For example, Internet victimisation risk factors, such as rule-breaking behaviour, mental health issues, and social isolation, are very similar to the risk factors for so many other youth behavioural problems [2-6].

Therefore, interventions aiming to promote digital literacy among children and young people may consider backing activities that have already been shown to reduce related risks factors [1]. While prevention and promotion interventions may have similar goals such as reducing cyberbullying or sexual exploitation, some important differences arise when focusing on the risks rather than on the opportunities that Internet can bring. Using the Internet can be a very healthy and rewarding activity as well as a potentially dangerous and unhealthy experience; it all depends on the user's awareness, knowledge and intentions.

Livingstone [7] suggests that risk, harm and vulnerability in children online can be researched by building on the literature for offline risk in children. Assessing risk and harm on the Internet, however, is particularly challenging because calculating the incidence rates of, for example, children being exposed to abuse online and the actual harm resulting from these hostile online encounters can be difficult. Indeed, there are no objectively verified and accurate statistics about how many children are exposed to inappropriate content, and therefore what is usually being reported is the 'risk of the risk' that might result in harm, which may be completely disproportionate as not all risk results in harm.

At present, the literature regarding online harm is sparse, making it difficult to understand whether a risk results in harm or how the Internet plays a role in known harm. Clearly, the situation regarding online risk is quite different from offline risk, however, it has been documented that children who are vulnerable offline are also more likely to be at risk online [8, 9]. Further understanding of the risk and protective factors that mediate the relationship between online and offline risk and harm seems mandatory, especially when considering a socio-technological context that is in constant change where the use of the Internet is widely spread among children and young people, creating new interactions between risk and protective factors.

For example, a recent systematic review of the effect of online communication and social media on young people's wellbeing [10] has showed contradictory evidence indicating that the Internet acts merely as a facilitator of human interaction and is itself value-free, neither promoting the good nor the bad. The findings from this review showed that online communication allow young people to increase the size and composition of their social networks can be either beneficial, because it can increase social support and social capital, or harmful through increased likelihood of exposure to abuse content or promotion of maladaptive coping strategies, such as self-harm [11]. Taking these findings into consideration, strategies to support the wellbeing of young people may wish to focus on the particular application being used, the communicative and non-communicative activities taking place, and the social support available offline to that individual to manage potential harm.

Due to the inevitable relation between humans and the digital world, it is more important than ever before that children and young people are familiar and confident with computers and technologies, not only because technology-related skills will optimise their future job opportunities, but also because promotes digital equalities and participation in society (e.g., digital citizenship) [12]. Therefore, it is vital that children are taught the benefits of new technologies and the associated risks but without frightening them or focusing too much on the risks associated with modern-day issues such as pornography, 'trolling', 'sexting', cyberbullying, and so on. For example, if we look back at previous research on youth prevention of substance misuse, we will find evidence showing that frightening messages do little to modify young people's risky or undesirable behaviour [13].

Recent evaluations and systematic reviews of Internet safety programmes showed that while participants can retain messages as indicated in follow-up questionnaires, there is little apparent impact on participant's behaviour [14-18]. There are several critical lessons to be learnt from previous research on prevention science that could guide new Internet safety educational materials. Recommendations include the development of interventions around strategies that are evidence-based and grounded in theory, meaning that the intervention explicitly defines why and how it is effective, indicating the social, behavioural and communication theories from which such strategies have been developed.

According to the literature [19, 20] effective prevention programmes target actual vs. perceived risks factors. For example, there is evidence to support that most young online sex crime victims are aware of the age difference of their perpetrator before meeting them face-to-face [21], therefore, educating young people about age deception is not as relevant as to provide education about judgement on sexual correspondence. Similarly, understanding risks and protective factors may help us understand who is actually vulnerable and avoid alarmist public perceptions that all children are 'at risk',

consequently increasing the media panic that results in demands to restrict children's Internet access, increase surveillance or violate data protection and online freedom.

Prevention programmes are most effective when they are integrated into school curricula, implemented consistently, and delivered by trained educators [22, 23]. Extracurricular activities, however, are often perceived as more flexible and dynamic than activities within the National Curriculum, which could prevent innovative activities from becoming a 'programme' ending up being bureaucratised and eventually fossilised. Understanding the relationship between young people and the Internet is crucial for designing effective interventions that promote not only the technical knowledge and skills necessary to successfully operate digital devices, but also promote a number of other aspects.

For instance, interventions could be designed to cover the cognitive and social skills necessary to recognise and integrate new models of social interaction (e.g., Facebook) and develop emotional intelligence to deal with the affective feedback from online interaction (e.g. Twitter). Interventions should also acknowledge alternative views and cultures and adapting to them (e.g., online forums), adjust self-control and self-awareness to manage time spent online (e.g., online gaming), recognise and address new types of malign intention (e.g., online grooming), adapt from a close, individual-based model of learning and creation to one based on collectively sourced collaboration (e.g., crowdsourcing), and so on. In this paper, the concept of *digital literacy* takes the humanities approach to consider the social skills and cultural competencies required to enabling participation within the new media culture.

According to Jenkins et al [24], there are three main problems that any digital literacy programme should address: the first issue tackles the inequalities in young people's access not only to new media technology and the Internet, but to skills and content that is most beneficial (i.e., what they call the participatory gap). The second issue focuses on the transparency problem or the potential commercial interests that may influence online decisions. This problem becomes apparent when analysing the advertising practices displayed on online gaming or the dangers of blending false or inaccurate information from facts. This is especially relevant when taking into consideration results from a systematic review on how children make sense of online resources showing a lack of both knowledge and interest in assessing how information was produced [25]. The third challenge focuses on the ethics, or how to encourage young people to become more reflective about the ethical choices they make online, and the potential impact on others. The ethics challenge is linked to digital citizenship and relates to the content young people post online, the content they access to (e.g., adult content), and compliance with implicit/explicit online community rules. These three issues (i.e., participatory gap, transparency and ethics) are central themes developed and dramatized in the iRights Youth Juries. These three problems related to the Right to Agency, the Right to Know and the Right to Digital Literacy described further below.

Finally, experts on prevention science [1] have also pointed out that creative and multi-faceted approaches involving peers, parents, teachers and the general public on either generic awareness campaigns or more specific/targeted training is also desirable.

3. IRIGHTS YOUTH JURIES

This section briefly describes the iRights Youth Juries, a new methodological approach for the promotion of digital literacy among children and young people. These juries take into consideration all the cumulative evidence and recommendations on online risk and protective factors, including the fuzzy links between risk, harm, and vulnerability, the need for a theoretical context, known predictors for successful prevention programmes such as implementation and delivery, the issues that literacy programmes should address, and who to involve on such programmes.

2.1Juries

This paper presents an innovative methodology to bring people together and facilitate reflection upon the issue of digital rights. What we are calling juries are similar to focus groups, but unlike many focus groups, juries have an explicit objective of arriving at clear recommendations regarding digital rights. Using the terminology of 'juries' is a important decision, as it is to be hoped that participants will subsequently feel a sense of responsibility as decision-makers, and facilitate participation and discussion.

How the jury is delivered and implemented is also extremely important, not only because the juries should be replicable and participants' outputs should not depend on the personal attributes of the facilitator or educator, but because explicit training, guidelines, and processes are in place, and a sense of ownership, responsibility, and care are also part of the training. For example, understanding the current evidence on online risks and protective factors is important to ensuring that accurate information and facts are discussed during the deliberation process.

It has been consistently shown that interactive programmes with skills training offered over multiple sessions outperform non-interactive, lecture-based, one-shot programmes [19, 26]. Currently, our juries are highly interactive and the scripts developed to dramatize the scenarios have been co-produced with young people to explore their personal concerns and online experiences. When co-producing scenarios with young people we are enhancing engagement opportunities, making these more real, easier to relate to, and consequently, maximising youth involvement on discussions.

The aim of our juries is not only to find out what participants (i.e. the "jurors") think and feel about the experiences of the digital world, but to discover what shapes their thinking and whether they are open to changing their minds in the light of discussion with peers or exposure to new information. In order to explore such questions, we are interested in discussing i) the reasons that jury members give for adopting particular perspectives and positions; and ii) the extent to which participant's perspectives and positions change, individually and collectively, between their arrival on the jury session and their departure. The jury session is typically lead by a trained facilitator, whose task is to provide a safe space for participants to express themselves freely and critically while demystifying issues around technology, data privacy, informed consent, and so on.

3.1 Vignettes

The use of dramatic scenarios builds upon the methodological research tradition of using *vignettes* as prompts to elicit reflective responses from participants. Vignettes are more frequently use in applied drama within educational settings which has a long tradition and for which there is extensive evidence on the underlying social, cognitive and emotional processes associated to applied drama for facilitating learning and development [27-29].

Bloor and Wood [30] define vignettes as: "A technique used in structured and in-depth interviews as well as focus groups, providing sketches of fictional (or fictionalized) scenarios. The respondent is then invited to imagine, drawing on her own experience, how the central character in the scenario will behave. Vignettes thus elicit situated data on individual or group values, beliefs and norms of behaviour. While in structured interviews respondents must choose from a multiple-choice menu of possible answers to a vignette, as used in in-depth interviews and focus groups, vignettes act as a stimulus to extended discussion of the scenario in question."(pp.183)

While the format of vignette presentation can vary including short video clip presentation and live acting, its aims and objectives are usually the same: to facilitate discussion, reflection, and deliberation amongst a group of young people (e.g. in this case, the jury) that may develop new attitudes, opinions, and interpretations about their digital rights and therefore, the potential benefit and harm associated with specific online activities. Vignettes can take several forms and their development and administration should always protect the research participants, especially when sensitive issues are being presented [31]. Usually vignettes are short stories that are read out loud to participants. Some researchers have used film and music, while others have used interactive web content or live acting, with its value deriving from combining the stimulus of the vignette method with the liveness and indeterminacy of the applied drama/theatre-in-education tradition.

The interpretation of responses to the scenarios entails complex analysis, involving the need to be clear about what we think responses represent, the extent to which there is a relationship between expressed beliefs and actions, the possibility that some participants might have felt under pressure to 'give the right answer', and the degree of consistency between post-scenario comments and broader findings from the group session tapes' and transcripts' [32, 33].

Vignettes have been used by researchers from a range of disciplines, including scholars studying public acceptance of mentally ill residents within a community [34], multicultural integration in neighbourhoods [35], the neglect and abuse of elderly people [36] and early onset dementia [37]. Vignettes have proved to be particularly useful in eliciting reflective responses from groups of young people: Barter and Renold [38] used them very successfully in their research with young people exploring violence in residential children's homes; Conrad [39] used vignettes as a way of talking to young rural Canadians about what they considered to be 'risky activity'; Yungblut et al [40] used them in their work with adolescent girls to explore their lived experiences of physical exercise; and Bradbury-Jones et al [41] employed vignettes to explore children's experiences of domestic abuse. To date we are not aware of any published research using vignettes to promote digital literacy.

3.3. iRights Youth Juries

This paper follows a series of iRights Youth Juries held in three UK cities including twelve young people per session aged 12-17 and from diverse socio-economic backgrounds. These juries illustrate the 'improvised drama' element of a piece of research lead by iRights [42], a new civil society initiative that is working to create a future where all young people have the fundamental right to access the digital world 'creatively, knowledgeably and fearlessly'. The juries were developed in collaboration with the SHM Foundation, The University of Leeds, and The University of Nottingham to explore five predefined digital rights and their implications with juries of young people. The following are the five digital rights covered:

1. The Right to Remove: 'Every child and young person under 18 should have the right to easily edit or delete any and all content they themselves have created. It must be right for under 18s to own content they have created, and to have an easy and clearly signposted way to retract, correct and dispute online data that refers to them.'
2. The Right to Know: 'Children and young people have the right to know who is holding or profiting from their information, what their information is being used for and whether it is being copied, sold or traded. It must be right that children and young people are only asked to hand over personal data when they have the capacity to understand they are doing so and what their decision means. It must be also be right that terms and conditions aimed at young people are written so that typical minors can easily understand them.'
3. The Right to Safety and Support: 'Children and young people should be confident that they will be protected from illegal practices and supported if confronted by troubling or upsetting scenarios online. It must be right that children and young people receive an age-appropriate, comparable level of adult protection, care and guidance in the online space as in the offline. And that all parties contribute to common safety and support frameworks easily accessible and understandable by young people.'
4. The Right to Make Informed and Conscious Decisions (The Right to Agency): 'Children and young people should be free to reach into creative and participatory places online, using digital technologies as tools, but at the same time have the capacity to disengage at will. It must be right that the commercial considerations used in designing software should be balanced against the needs and requirements of children and young people to engage and disengage during a developmentally sensitive period of their lives. It must also be right that safety software does not needlessly restrict access to the Internet's creative potential.'
5. The Right to Digital Literacy: 'To access the knowledge that the Internet can deliver, children and young people need to be taught the skills to use and critique digital technologies, and given the tools to negotiate changing social norms. Children and young people should have the right to learn how to be digital makers as well as intelligent consumers, to critically understand the structures and syntax of the digital world, and to be confident in managing new social norms. To be a 21st century citizen, children and young people need digital capital.'

During the iRights Youth Juries, participants put the Internet on trial by deliberating on a series of real-life digital scenarios, previously produced in partnership with young people and brought to life by live actors. To work in equal partnership with children and young people is relevant to further develop the iRights Youth Juries and ensure vignettes present real issues and experiences to which young people can relate to and maximise their ecological validity. Working with young people as equal partners is also important to guarantee that the language used to dramatize the scenarios resonates with their vocabulary and expressions. Because scenarios have to be co-produced with local young people, vignettes are idiosyncratic and sensitive to cultural differences as they should represent a specific and distinct point in time, avoiding universalistic terms. In this way, the scenarios developed for this first wave of iRights Youth Juries will differ from those developed in the

near future as smart phone applications, computer games and lexicon around technologies rapidly evolve with time.

In relation to the three main problems outline by Jenkins et al., (i.e., participatory gap, transparency and ethics) our juries have been designed to promote social skills and cultural competencies through dialogue, collaboration, and discussion. The juries offer objective information about data privacy issues and a space for reflection to develop critical-analysis skills on how media shapes perceptions of the word. The dilemmas or conflicts that the scenarios bring to life include an element of reflection on the negative as well as the positives exhibited on the Internet. These dilemmas also encourage young people to pull knowledge and reconcile conflicting information to form a coherent picture. This is a form of problem solving valuable in shaping all kind of relationships (e.g., knowledge, community, tools, etc.).

The presence of live actors added a realistic dimension to the deliberation process and served to highlight key themes and issues by bringing them to life and stimulate discussions. This could be considered a form of simulation, encouraging young people to interpret and construct models of real-world processes. As the dramatized scenarios are highly dynamic, allowing space for improvisation and interaction between actors and participants, young people can formulate hypotheses of 'what is going to happen next', test different variables in real time, and modify or refine their interpretation of the 'real world' while engaging them in a process of modelling (i.e., learning that takes place in a social context through observation). It is well known [43, 44] that students learn more through direct observation and experimentation that simply by reading text books, or listening in the classroom setting. Simulations not only broaden the kinds of experiences students may have but brings capacities to understand problems form multiple perspectives, to assimilate and respond to new information.

These juries are embedded in a research process designed to explore digital rights and their implications with juries of young people. Specifically, the research project has been designed to capture reflections on (1) their experiences of anxiety, uncertainty, frustration, and aspiration in using digital technologies; (2) their understanding of who 'runs' the Internet, who polices it, what 'it' is, and how far they feel they can control their digital experiences; (3) their sense of their own digital literacy and its limitations; (4) their responses to new information about the Internet and digital technologies; (5) the relevance and effectiveness of specific digital rights (see below) in relation to such experience; (6) appropriate language and techniques for sharing and disseminating digital rights; and (7) ways of further engaging young people in thinking about and acting upon their rights as digital citizens.

Future youth jury developments should incorporate skills training over multiple sessions. For example, if a scenario focuses on the 'right to know', a more hands-on session or workshop could focus on how to avoid third-party tracking cookies designed to compile long-term records of individual's browsing histories. Skills training could complement the deliberation process on potential privacy concerns that cookies represent when storing passwords and sensitive information, such as credit card numbers and address. Ideally, juries should be offered on more than one session and present a repertoire of scenarios that have been co-produced with a local representative sample of children and young people to illustrate up-to-date and culturally relevant online youth concerns and celebrations. The core measures used within the current study included semi-structured interviews and questionnaires completed before and after the jury, designed to assess attitudinal changes.Our current research focuses on comparing iRights Youth Juries' outcome measures (i.e., attitudinal change and semi-structured interviews) when, instead of live acting, short video clips are presented. While live acting adds an element of excitement, its high costs and complex logistics may impede wider dissemination and consequently minimise participation. Video is a plausible format for secondary schools where iRights Youth Juries can be easily recreated and delivered within both drama and IT school departments. During ETHICOMP2015 we intend to explore conference attendees' rationales for accepting and rejecting accounts of social reality or proposals for digital strategies or policies (e.g. online data protection).

We suggest initiating this session by allocating time for delegates to speak freely about which digital rights should be considered and their experiences of digital activity. This can be done in small groups to ensure all voices are heard. The jury can vote on the digital rights proposed in each group and the three that received the most votes could be selected for further deliberation. Each stage of the jury deliberation will conclude with a facilitated discussion in which participants are urged to formulate one key principle that would allow them to experience greater control over the aspect of digital activity for which the digital rights were under consideration. During each of these discussions jury participants witness a scenario: a short video clip of an incident or dilemma presented with a view to eliciting thoughtful resolutions from participants. Participants are encouraged to discuss each of the scenarios or vignettes and decide how they think the dramatized situation should be resolved. Resolutions and their consequences are then discussed further.

This session is part of conference track 'New ideas on bringing people together / novel formats', and these are some of the prompts or topics ETHICOMPT2015 delegates may reflect on and offer advice relating to:

- potential and possible digital rights
- the relevance and effectiveness of digital rights
- the ways in which digital rights (or their absence) can affect us
- techniques for sharing and disseminating digital rights
- ways of further engaging with the general population in thinking about and acting upon digital rights

This method of deliberation – space for participants to express, compare and make sense of their views and experiences - is expected to generate thoughts among delegates for critical and reflective thinking about digital rights with the view to modify undesirable behavior. We believe iRights Youth Juries will bring an engaging and exciting element to ETHICOMP2015, and in the near future an alternative to existing Internet Safety programmes offered to school and parents that risk lacking relevance to members of the cohort for whom they are designed.

4. ACKNOWLEDGEMENTS

This work forms part of the Citizen-centric Approaches to Social Media Analysis (CaSMa) project, supported by ESRC grant ES/M00161X/1 and based at the Horizon Digital Economy Research Institute, University of Nottingham. For more information about the CaSMa project, please see http://casma.wp.horizon.ac.uk/.

5. REFERENCES

[1] Jones L. M. 2010. The future of Internet safety education: critical lessons from four decades of youth drug abuse prevention. *Pubios Project.* http://publius.cc/

[2] Stice, E., Shaw, H., Bohon, C., Marti, C. N., and Rohde, P. 2009. A Meta-Analytic Review of Depression Prevention Programs for Children and Adolescents: Factors That Predict Magnitude of Intervention Effects. *Journal of Consulting and Clinical Psychology*, 77(3), 486-503. doi: Doi 10.1037/A0015168

[3] Skiba, D., Monroe, J., and Wodarski, J. S. 2004. Adolescent substance use: Reviewing the effectiveness of prevention strategies. *Social Work*, 49(3), 343-353.

[4] Durlak, J. A., and Wells, A. M. 1997. Primary prevention mental health programs: The future is exciting. *American Journal of Community Psychology*, 25(2), 233-243. doi: Doi 10.1023/A:1024674631189

[5] Wells, M., and Mitchell, K. J. 2008. How do high-risk youth use the Internet? Characteristics and implications for prevention. *Child Maltreatment*, 13(3), 227-234. doi: Doi 10.1177/1077559507312962

[6] Durlak, J. A. 1998. Common risk and protective factors in successful prevention programs. American Journal of Orthopsychiatry, 68(4), 512-520. doi: Doi 10.1037/H0080360

[7] Livingstone, S. 2010. E-Youth:(future) policy implications: reflections on online risk, harm and vulnerability. In *Proceedings at the e-Youth: balancing between opportunities and risk* (Antwerp, Belgium 27-28 May 2010) UCSIA & MIOS http://eprints.lse.ac.uk/27849

[8] Livingstone, S., Haddon, L., and Görzig, A. (Eds.) 2012. *Children, Risk and Safety Online: Research and policy challenges in comparative perspective.* Bristol: The Policy Press.

[9] Bradbrook, G., Alvi, I., Fisher, J., Lloyd, H., Moore, R., Thompson, V., et al. 2008. *Meeting Their Potential: The Role of Education and Technology in Overcoming Disadvantage and Disaffection in Young People*. Coventry: Becta.

[10] Best, P., Manktelow, R., and Taylor, B. 2014. Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review*, 41, 27-36. doi: DOI 10.1016/j.childyouth.2014.03.001

[11] Duggan, M., and Brenner, J. 2013. *The demographics of social media users.* Pew Internet Research Centre's Internet and American Life Project (http://pewInternet.org/Reports/ 2013/Social-media-users.aspx).

[12] Mossberger, K., Tolbert, C., McNeal, R. 2007. Digital Citizenship. MIT Press.

[13] Lynam, D. R., Milich, R., Zimmerman, R., Novak, S. P., Logan, T. K., Martin, C., . . . Clayton, R. 1999. Project DARE: No effects at 10-year follow-up. *Journal of Consulting and Clinical Psychology*, 67(4), 590-593.

[14] Dresler-Hawke, E., and Whitehead, D. 2009. The Behavioral Ecological Model as a Framework for School-Based Anti-Bullying Health Promotion Interventions. *Journal of School Nursing*, 25(3), 105-204. doi: Doi 10.1177/1059840509334364

[15] Chibnall S, Wallace M, Leicht C, Lunghofer L. 2006.*I-SAFE Evaluation: Final Report.* Fairfax, Virginia: Caliber.

[16] Brookshire M, Maulhardt C. 2005 *Evaluation of the effectiveness of the NetSmartz Program: A study of Maine public schools*. Washington DC: The George Washington University.

[17] Crombie G, Trinneer A. 2003. Children and Internet Safety: An Evaluation of the Missing Program. A Report to the Research and Evaluation Section of the National Crime Preventure Centre of Justice Canada. Ottowa: University of Ottowa.

[18] Shillair, R., Cotten, S. R., Tsai, H. Y. S., Alhabash, S., LaRose, R., and Rifon, N. J. 2015. Online safety begins with you and me: Convincing Internet users to protect, themselves. *Computers in Human Behavior*, 48, 199-207. doi: DOI 10.1016/j.chb.2015.01.046

[19] Bond LA, Carmola Hauf AM. 2004. Taking stock and putting stock in primary prevention: Characteristics of effective programs. The Journal of Primary Prevention, 24(3):199-221.

[20] Winters K.C., Fawkes T., Fahnhorst T., Botzet A., Augsut G. 2007 A synthesis review of exemplary drug abuse prevention programs in the united states. *Journal of Substance Abuse Treatment*, 32:371-380. doi: DOI 10.1016/j.jsat.2006.10.002

[21] Wolak, J., Finkelhor, D., Mitchell, K. J., & Ybarra, M. L. 2008. Online "Predators" and their victims - Myths, realities, and implications for prevention and treatment. *American Psychologist*, 63(2), 111-128. doi: Doi 10.1037/0003-066x.63.2.111

[22] Durlak, J. A., and DuPre, E. P. 2008. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350. doi: DOI 10.1007/s10464-008-9165-0

[23] Payne, A.A., Eckert, R. 2009. The relative importance of provider, program, school, and community predictors of the implementation of quality of school-based prevention programs. Society for Prevention Research. 1-16.

[24] Jenkins, H., Purushotma, R., Weigel, M., Clinton, K. and Robison, 2009. *A Confronting the challenges of participatory culture. Media education for the 21st Century*. MIT Press.

[25] Buckingham, D. 2005. The Media Literacy of Children and Young People: A Review of the Literature. Centre for the Study of Children, Youth and Media, Institute of Education, University of London, 22. http://www.ofcom.org.uk/advice/media_literacy/medlitpub/ medlitpubrss/ml_children.pdf (accessed September 2006).

[26] Ennett ST, Tobler NS, Ringwalt CL, Flewelling RL. 1994How effective is drug abuse resistance education? A metaanalysis of project DARE outcome evaluations. *American Journal of Public Health*, 84(9):1394-1401.

[27] Tombak, A. 2014. Importance Of Drama In Pre-School Education. 3rd Cyprus International Conference on Educational Research (Cy-Icer 2014), 143, 372-378. doi: DOI 10.1016/j.sbspro.2014.07.497

[28] Winston, J. 2015. Imagining the real: towards a new theory of drama in education. *British Journal of Educational Studies*, 63(2), 252-254. doi: Doi 10.1080/00071005.2015.1035913

[29] Burn, A. (2011). Drama Education with Digital Technology. English in Education, 45(1), 104-106. doi: DOI 10.1111/j.1754-8845.2010.01086.x

[30] Bloor, M. and Wood, F. 2006. *Keywords in Qualitative Methods: A Vocabulary of Research Concept.* Sage Publications Ltd. London.

[31] Bradbury-Jones, C., Taylor, J., and Herber, O. R. 2014. Vignette development and administration: a framework for protecting research participants. *International Journal of Social Research Methodology,* 17(4), 427-440. doi: Doi 10.1080/13645579.2012.750833

[32] Barter, C., and Renold, R. 1999. The use of vignettes in qualitative research. *Social Research Update*, 25. http://sru.soc.surrey.ac.uk/SRU25.html

[33] Finch, J. 1987. The vignette technique in survey research. *Sociology*, 21(2), 105-114.

[34] Aubry, T. D., B. Tefft, and R .F. Currie. 1995. Public-Attitudes and Intentions Regarding Tenants of Community Mental-Health Residences Who Are Neighbours. *Community Mental Health Journal,* 31(1): p. 39-52. Doi 10.1007/Bf02188979

[35] Schuman, H. and L. Bobo. 1998. Survey-Based Experiments on White Racial-Attitudes toward Residential Integration. *American Journal of Sociology*, 94(2): p. 273-299. Doi 10.1086/228992

[36] Rahman, N. 1996. Caregivers' sensitivity to conflict: The use of the vignette methodology. *Journal of Elder Abuse and Neglect, 8*(1), 35-47. Doi 10.1300/J084v08n01_02

[37] Smythe, A., Bentham, P., Jenkins, C., and Oyebode, J. R. 2015. The experiences of staff in a specialist mental health service in relation to development of skills for the provision of person centred care for people with dementia. *Dementia-International Journal of Social Research and Practice*, 14(2), 184-198. doi: Doi 10.1177/1471301213494517

[38] Barter, C., and Renold, E. 2000. 'I wanna tell you a story': Exploring the application of vignettes in qualitative research with children and young people. *International Journal of Social Research Methodology,* 3(4), 307-323.

[39] Conrad, M. 2004. Family life and sociability in upper and lower Canada, 1780-1870. A view from diaries and family correspondence. *Revue D Histoire De L Amerique Francaise*, 58(1), 124-126.

[40] Yungblut, H. E., Schinke, R. J., and McGannon, K. R. 2012. Views of adolescent female youth on physical activity during early adolescence. *Journal of Sports Science and Medicine*, 11(1), 39-50.

[41] Bradbury-Jones, C., Taylor, J., Kroll, T., and Duncan, F. 2014. Domestic abuse awareness and recognition among primary healthcare professionals and abused women: a qualitative investigation. *Journal of Clinical Nursing*, 23(21-22), 3057-3068. doi: Doi 10.1111/Jocn.12534

[42] www.iRights.uk

[43] Gaba D. *Human work environment and simulators*. In: Miller RD, editor. In Anaesthesia. 5th Edition. Churchill Livingstone: 1999. pp. 18–26.

Alterity and freedom of information on the Internet – The loss of Net Neutrality in contemporary literature

Jasmin Hammon
University of Augsburg / University of Limoges
Phd student (supervisors: Professor Rotraud von Kulessa, Professor Till Kuhnle)
Research group EHIC (Espaces humains et Interactions Culturelles)
31, rue François Perrin
87000 Limoges
0033 78 27 52 452
jasmin.hammon@etu.unilim.fr

ABSTRACT
Giving a short overview of the technical innovation of Deep Packet Inspection and "Internet fast-lanes", this paper shows the ethical dimension of giving up the founding principle of the open Internet. These new means of network management are fiercely discussed because of their inherent threat of censorship and attack on privacy. This paper will attempt to explain why the experience of otherness and both the freedom of information and the freedom of expression are endangered if Net Neutrality is no longer protected. Furthermore, it shows by means of some representative examples how this subject is reflected in contemporary literature.

Categories and Subject Descriptors
C.2.3 [**Computer-Communication Networks**]: Network Operations: Network management

General Terms
Human Factors, Economics, Legal Aspects, Security.

Keywords
Net Neutrality, Open Internet, Internet "fast-lanes", prioritization, network management, broadband, bandwidth, network capacity, freedom of information, freedom of expression, agora, censorship, privacy, democracy, alterity, contemporary literature, European Union, United States of America, Canada, best-effort delivery.

1. INTRODUCTION

Contemporary literature is always considered as a symptom of Zeitgeist, discussing in more or less factually correct or fictionally alienated way current subjects of general interest. Or as Frank Kermode emphasizes, fictions "correspond to a basic human need, they must make sense, give comfort" [1]. Sometimes, literature envisions future scenarios, for instance in utopian or dystopian subgenres. It is one important function of literature to help imagine the consequences of our actions which we cannot estimate for the moment because of lack of experience. German philosopher Günther Anders explains how mankind is unable to foresee the effects of new technology [2]. Even if he is referring to nuclear technology, his theory of mankind's lack of imaginative power can be, more globally, associated with all technologies. So, in literature we try out all possibilities – it therefore helps to discuss social and political subjects, as well as technological progress and its ethical dimension. That is why the freedom of information and the freedom of expression are important subjects in contemporary literature. The debate of an Open Internet influences the public and private sphere of people as citizens, consumers and private persons, it affects the working world as well as leisure time and social interactions. The second chapter of this paper demonstrates some of the major violations of power by Internet service providers (ISP). These abuses, like blocking political sites or censoring live streams, fuel the already grim debate over Net Neutrality and independent information. This subject also is argued in contemporary literature as we show in the last chapter of this paper.

This paper attempts to answer the question why there is this profound criticism and fear regarding so called "Internet fast-lanes". It tries to explain why network management and prioritization might lead to the loss of the freedom of information and the freedom of expression and how censorship would inhibit the experience of alterity. The first section gives a short overview of the latest news and the subject's evolution in European and American news and highlights the global significance of Net Neutrality. Section 1.2. and 2. will demonstrate how the new technologies manipulate the access to information and therefore influence the experience of alterity which is considered as crucial for developing democratic and critical opinions. The third part refers to examples of contemporary literature where the importance of independent information and free expression reflect real fears.

1.1 Current status of Net Neutrality in Europe, Canada and the USA

After the historic vote of the European Parliament for a stronger protection of the free Internet in April 3rd 2014, the topic "Net Neutrality" disappeared largely off the daily news. Even after June 30th 2015, when the same institutions suddenly back pedal and trade the gained Net Neutrality policy against cheaper roaming fees, the reactions in the press had been relatively restrained. One of the few articles published was in the German (online) newspaper "Die Zeit" which summarizes the new proposition which allows ISP to create "Internet fast-lanes" to provide specialized services, on condition that they can ensure a minimum of quality for the "normal" Internet access. The formerly strict clauses concerning Net Neutrality have been watered down to vague expressions [3] and speculations of specialized services like automotive cars or telemedicine, which are still mere "phantoms" than reality[4], and the term "Net Neutrality" was replaced by "open internet access" - which suggests the possibility of a two class Internet.

Therefore, the equality of data is still at stake. The German Government claims, since the beginning of the debate, the opening to prioritization, also called Internet "fast-lanes" to progress data traffic management. Angela Merkel wants to improve the German network, which is one of the slowest and most obsolete networks in Europe. However, this is not to be funded with state subsidies, but it is by means of capitalism, namely the sale of Internet fast-lanes, that she wants to upgrade the broadband net. While the German Government does not feel responsible for covering the costs connected to the expensive maintenance and modernization of the network, the German population is equally reluctant to raise the funds on their own. Therefore, offering online services priority in data traffic could be a lucrative solution. Nevertheless, critics fear the loss of an uncensored Internet. That is the reason why Net Neutrality is currently, even after the trilogue's decision, a controversial issue, especially in Europe, Canada and the USA. Surprisingly, the issue appears to be given less media attention in European countries or it has not reached public awareness yet.

The Federal Communications Commission (FCC) of the USA decided in February 2015 to protect Net Neutrality in order to provide free access to information and exchange of ideas. The ISP "must act in "public interest" and should be regulated the same as telephone and cable providers" [5]. This means, that they have to grant access to their networks even to competing companies to ensure a fair competition. The FCC's decision also influenced the Canadian Radio-television and Telecommunications Commission (CRTC), which "always regulated the Internet as a utility, which has made it somewhat easier for the CTRC to step in" [6]. The situation in Canada, however, is different, because the networks produce overage, so the ISP can provide "wholesale telecommunications services to competitors" [7]. To avoid dominant services and companies to accumulate too much of the data traffic, Canada decided on strict Net Neutrality policies [8]. Both, the USA and Canada, are economically strong partners of the European Union, being both in negotiations for transatlantic treaties (TTIP, resp. CETA), they have immense effects on European decisions.

In French online newspapers, such as www.lemonde.fr, even providing a special section for Net Neutrality [9], or www.lefigaro.fr, Net Neutrality is rarely mentioned [10], or, the access to articles is fee-based [11]. In addition, the French Government supports the interests of the big telecom operators [12]. Still, there is "La Quadrature du Net", a very active French advocacy group fighting for digital rights and against prioritization. The way the issue of Net Neutrality is being addressed varies strongly from one European country to the next; however, some of the governments' plans are in stark contrast to the needs and wishes of the general population, as shown with the example of Germany and France.

1.2 Technical and legal details of Net Neutrality

There is this perhaps old-fashioned utopia of having an Internet in which each bit is treated equally, the so-called "best-effort delivery", where all data packages are strictly delivered at the same speed and one after the other, "first-in-first-out" [13]. The network capacity, however, today is almost exhausted and the mobile Internet bandwidth is physically limited [14]. Despite the imminent limit of capacity, the market is still growing. In 2011, more than 2,6 billion gigabyte had been sent through the German networks and experts estimate this to increase by a factor of 20 until 2020 [15]. The same development can be assumed for all European, American and Canadian networks. The collapse of the networks seems predestined when we continue to transfer each bit equally.

Basically, there are two major solutions to the problem: Either, the networks are extended, or the existing networks are managed in a more efficient way. The idea is to create "Internet fast-lanes" for urgent data transmission, such as live streams, telemedicine ("e-health services"), communication via Voice over Internet Protocol (VoIP). These services need real time transmission whereas an email can wait to be delivered. This "prioritization" would be charged differently by the ISP: more speed, more fees. The ISP have been for a while discontent, because some popular online services use most of the bandwidth of the network and slow down the data traffic. YouTube and Google occupied around 27% of the data traffic in 2008 in the U.S. [16], with rising tendency, and the ISP claim to share the profits or to be compensated for the domination of their networks. Just like medieval road tolls, the most frequented companies like Amazon, Facebook, Google or YouTube would pay more for using most of the data traffic.

Prioritization, or "fast-lanes", seems to be a perfect solution for the online services and the ISP and even the clients would benefit from faster data transmission, especially when using their favorite services. Furthermore, the ISP would make enough profit to maintain a quick and stable Internet for their clients. The negative effect of prioritization is the so-called "posteriorization" [17], the discrimination of data, which could influence and distort economic competition, communication and the access to information on the Internet. It could provoke monopoles of opinions and censorship exercised by private companies [18]. Critics fear that only the already dominant companies have the financial funds to reserve the bigger part of the data traffic and smaller companies as well as start-ups would be discriminated. Consequently, those could only transfer in slower quality. The dominant services would have a direct influence on which information users would get. These dangers are even more menacing when there is no legislation to control the market and the service providers. Thus, the governments would hand over responsibility entirely to private companies. Therefore fast-lanes would handicap innovation and lead to a monopolization of the market. In the USA, the economic leaders argue that a healthy

capitalism regulates itself. The consumers directly affect success or failure of services and companies, so there is a constant change on the market which prevents by itself any monopolization. There is even a certain fear of having a "socialism of the Internet" when applying Net Neutrality [19]. In addition, many critics of Net Neutrality worry about excessive governmental influence on economy, privacy and the Internet [20]. On the contrary, "the guiding principle of net neutrality is to preserve the freedom and openness online that consumers currently enjoy, not to dictate what constitutes permissible content and expression online." [21].

Another technical innovation in the context of Net Neutrality threatens the Open Internet: the deep packet inspection (DPI), which is a tool of network management, through which an ISP can read the content of a data package in order to decide both the way and the speed it is routed. Very urgent data packages would be transferred first. So far, only the header has been read. With DPI, critics fear the violation of privacy and freedom of information, because DPI acts on all levels from 2 to 7 of the data package (levels of OSI reference model) [22], in which the content of the data package is deposited. Abusers of this technology could read the concrete wordings of emails or commit other violations of privacy. This contradicts Article 12 of the Universal Declaration of Human Rights (UDHR) embodying the privacy of someone's home and communication [23]. In Europe, deputies asked for DPI measures during the Net Neutrality debate, hoping to enhance the pursuit of child pornography [24] and, less severe, copyright infringements [25].

"Fast-lanes" and DPI also offend Article 19 of the UDHR securing the freedom of opinion and information:

"Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers." [26]

The next section will explain more precisely how Net Neutrality ensures this freedoms and democracy.

2. DEMOCRACY, ALTERITY AND THE OPEN INTERNET

The most threatening danger of such a two class system of the Internet would be to end the Internet's most important and most democratic value: the freedom of information and speech. Service providers and operators would not only block their rivals by buying most of the data traffic, but they could also influence the information passing through to their clients. Around 2005, several operators already have abused their control and that is how the subject "Net Neutrality" suddenly became important in the press. Legally questionable measures [27] like "Canadian Internet access provider Telus [has been] blocking access to their worker's labor union website during a lockout" [28] showed the power of operators over the content of and the access to Internet as well as the influence on political processes.

Other telecommunication companies such as American Telco [29] or German T-Mobile [30] also abused their power to close access to competing services like VoIP communication in favor of their own supply. VoIP uses Internet instead of ISDN like "classic" telecommunication and is less expensive, or even free of charge, compared to "classic" telephony, especially when users communicate with interlocutors abroad. When VoIP software Skype was introduced for iPhones in Germany in 2009, German telecommunication provider T-Mobile inhibited the new service. In official statements of the company, the press spokesman Alexander von Schmettow, declared, that this measure aims at ensuring the high quality of the network, expressing worries about provoking network congestion of the mobile [31]. Clients and critics like Macnotes, however, suspect other reasons, arguing, that VoIP is cheap and in direct concurrence to the telecommunication services of T-Mobile [32]. The German Federal Network Agency intervened and forced T-Mobile to either unblock VoIP or offer the VoIP with an additional fee [33]. Not only T-Mobile, but 23% of European ISP had imposed restrictions on VoIP at that time, underlines Bortnikov who analyzes the legal basis of Net Neutrality in Germany. He shows the difficulty of defining Net Neutrality by German law, because on the one hand, Net Neutrality is a subject of German Telecommunication Law and therefore ISP who often are at the same time (in the case of T-Mobile) the owners of the telecommunication and Internet networks, also have to open their networks to competing companies. They are not allowed to discriminate competitors with lower quality or even blocking access [34]. So in Germany, as generally in the European Union, governments ensure a minimum of quality and even prohibit any discrimination due to Article 22, paragraph 3, of the Universal Service Directive [35] as well as the European Competition Law [36]. On the other hand, this Directive would not be violated when ISP create a second, faster, prioritized Internet for clients willing to pay more [37]. As shown earlier, the newest European decisions are rather vague, so users and clients again have to fear more restrictions of the Open Internet. In the United States of America, the status of the ISP has always been in discussion, if Internet is considered being an information service or being a telecommunication service [38]. In 2007, when multinational telecommunications corporation AT&T blocked the sound during a concert stream of the band Pearl Jam, because of lyrics criticizing former US president Bush [39], the American Government had to redefine Internet and Net. After years, the FCC decided in February 2015 to protect Net Neutrality, after having classified the Internet as telecommunication service. Now, it falls under the 1934 legislation, so the Government has more influence on the ISP in demanding undiscriminated data transmission [40]. This measure guarantees the American citizens to access to independent information and it leaves the control over the Internet in "the hands of the public [41].

Net Neutrality has also always been an ethical subject, considering the experience of alterity in order to form enlightened and democratic opinions. The presented examples of private companies influencing their client's access to independent information and communication underline the necessity of an Open Internet. They also prove the Internet to already being subjected to limitation and censorship and explain why users disagree with further restrictions. As Danny Kimball analyzes the evolution of the term "Net Neutrality", the ISP argue "technical efficiency, marketplace competition" [42], whereas users claim "freedom of expression" [43]. Even more, the users want to regain the freedom they had in the early times of Internet when every data package was treated equally and access to information and exchange were not inhibited or discriminated. Fabien Granjon emphasizes the importance of online blogs for expressing and exchanging opinions and testimonies independently from the "classic" media and official political statements:

"Online mobilizations have always been presented as spheres of expression to fight for reappropriation of the public debate […] Internet has also often been presented as technical support of speech and mass distribution to the largest audience, beyond the usual spheres of exercising citizenship. […] [blogs] offer precisely this form of producing civil information by independent individuals (who are not active in political parties or traditional associations)" [44].

That is why Internet often is, for instance by jurist Mario Martini in his inaugural lecture, described as modern agora because of its neutral and open access to information. Internets benefit is its unlimited possibilities of communication, because it bundles all common media [45]:

"It has become one of the most important public spheres of social interaction and democratic participation, just like a virtual agora." [46]

One should not forget the direct political influence of the Internet. Today, it is easy to start online petitions against for instance questionable laws or against economic and social injustice - "everyone of those fights will be won or lost on the internet, which make the fight for the net the most foundational fight, if not the most important." [47]. This enthusiastic view on the importance of an Open Internet may seem naive, but it shows the user's need for free and independent information and debate.

Not only in politics, but also in economy, Internet has proven, with its openness and undiscriminating transmission of all data, to be a pillar for creativity and company start-ups [48]. The idealized old-fashioned Internet has per se no controlling organs, no hierarchy, but it helps to connect and to interact, that is why it is a decentralized, egalitarian pool of content and information and this is the basis of i's success [49], as Martini underlines the importance of an Open Internet.

On the social side, Internet is a democratic organ where users can exchange their points of view and political ideas or get into contact with different cultures and it is a place where they can find independent information.

Inernet is independent from institutional influence and offers participation in collective processes of creation, communication and opinion making and it even can catalyze the democratization and open access to information and knowledge. Martini clearly defends the principle of an unlimited Internet and recalls its crucial role during the Arab Spring and compares Internet with the political journals during the German Vormärz of the 19th century [50].

Christina C. Constantopoulou emphasizes the two "indispensable conditions" for democracy which are both provided online: "the freedom of expression" and "the uncensored circulation of ideas", independent of governments and also capitalism [51].

As we will show later, contemporary literature criticizes the loss of these freedoms, like in Margaret Atwood's "MaddAddam" trilogy, where private corporations literally "own" media and Internet.

The openness of the Internet expands not only on the access to information, but also on the equality of each member, the possibility for everyone to participate. As Constantopoulou explains, every user has the same rights to create or to join blogs which are not as filtered as the written press which often favors the opinions of experts, whereas the blog can be written by anyone, which creates a much more democratic exchange of ideas than the conventional press [52].

Consequently, online resources can be qualified authentic, direct, unfiltered and public. This is the potential to experience alterity which makes the Internet so valuable to us. Alterity, the "otherness", is everything different from our own culture, language and concepts. Online travel and food blogs can offer the users a glimpse on foreign countries and the customs of other people around the world. With services like Google Earth, we can "visit" places all over the planet without leaving our own house.

On YouTube, users share their experiences in encountering other ethnic or religious groups and discuss subjects like racism and discrimination [53]. Confronting alterity means dealing with the different and the strange and invites to take another perspective on the own convictions. Internet, with its endless means of information and personal exchange, offers the possibility to deal with this alterity and helps us to learn new concepts and ideas. In the experience of alterity, we discover new worlds, which deepens our education, broadens our own cultural horizon and helps developing our points of view in a democratic way. Therefore, facing and dealing with otherness is the basis of tolerance and acceptance, values which become more and more important in today's society with its growing radical tendencies. A "managed", prioritized, Internet and violations of privacy with technologies like DIP can give way to censorship and the fear of abuse persists and frightens Internet users. The new technologies of network management and network security could allow governments to filter and block Internet content, which would inhibit the users to access information and develop democratic opinions. Furthermore, people fear to become "transparent citizens", even more than they already are, in addition to the power of private companies, blocking all alternative concepts and information, the worries that they abuse their power to inhibit a fair and free competition – this is the core of the Net Neutrality debate. This is the main claim to maintain this alternative, identity creating sphere of exchange, where users form their ideas, where they encounter alterity to understand [54].

3. THE LOSS OF THE OPEN INTERNET IN CONTEMPORARY LITERATURE

Literature often reflects common fears and hopes, it can thus be seen as an indicator for social development, or as Jean-Paul Engélibert expresses it, a historic change happens before we are able to think about it [55]. Other way round, literature can develop thoughts and scenarios of a near or far future and therefore influences society.

The loss of free media and Internet is a common subject in this visionary fiction and the next paragraphs will present some significant examples. We have chosen the "Wool" trilogy by American author Hugh Howey, the "MaddAddam" trilogy by Canadian literature professor Margaret Atwood, and "La possibilité d'une île" by the guarantor of controversy "par excellence", Michel Houellebecq. They are only a small choice of the thousands of international bestsellers, but they still can serve as examples for this paper. As explained earlier, the European

Union, the United States and Canada are the most important economic driving forces discussing Net Neutrality at the moment. Their decisions on this subject have a major impact on the global evolution of the Internet. That is why literary examples of these territories have been chosen for the analysis.

The novels have in common to be extraordinarily famous and international bestsellers, so a certain interdependence between fiction and society can be assumed. Furthermore, they have (post) apocalyptic plots and question the future of mankind and the relations between human beings. Media and communication have a key role as means of interpersonal exchange and hence are a battlefield for ethics and democracy.

3.1 "Wool" trilogy by Hugh Howey

In Hugh Howey's bestselling "Wool" trilogy, for instance, the act of sending emails is strongly limited by censorship so that the people remain ignorant of the world of lies they live in. In his vision of a far future, a small part of mankind survives a nuclear war between the USA and his enemies, such as North Korea and Iran. The political leaders had been prepared for the worst case scenario and build fifty underground silos to safely hide a few thousands of people. The silos are self-sufficient and produce all they need inside of the facilities. Access to the contaminated outside is strictly forbidden and even an openly expressed wish to go outside is sentenced with death penalty. Female character Allison wants to convince her husband, sheriff Holston, that

"'No. I found the programs they use. The ones that make pictures on the screens that look so real,' She looked back to the quickening dusk. 'IT,' she said. 'Eye. Tee. They're the ones. They know. It's a secret that only they know." [56]

"Allison nodded. 'Expressing any desire to leave. Yes. The great offence. Don't you see why? Why is that so forbidden? Because all the uprisings started with that desire" [57].

These are measures to avoid uprisings and mass breakouts which would end in the death of thousands when the airlocks were opened and the contaminated outside air would flood into the silos. That is why the leading silo 1 blocks any contact between the silos, so the inhabitants do not know of their existence to prevent any attempts to go reach surface. The open exchange of ideas and the memories of the old world prove to be dangerous, seducing the silo inhabitants to risk their lives instead of being the germ cell of a new mankind. Therefore, the leaders of silo 1 want them to have limited and regulated media and communication systems, for example an email is more expensive than sending a messenger. One of the main characters of the trilogy is the new sheriff Juliette, a young woman working in Mechanics of silo 18, who discovers these political strategies and secretly consults her friend Peter:

"[Juliette:] 'Can you think of why it's cheaper to porter a paper note to someone than it is to just wire them from a computer?' […] [Juliette:] 'but you'd think we could all send and receive as many wires as we wanted.' […] [Juliette:] 'But what if it's for a different reason? What if someone made it expensive on purpose?' 'What? To make money?' Peter snapped his fingers. 'To keep the porters employed with running notes!' Juliette shook her head. 'No, what if it's to make conversing with each other more difficult? Or at least costly. You know, separate us, make us keep our thoughts to ourselves." [58]

The porters are not to be trusted either, having too much influence on the politics of the silo in keeping or giving away the secrets of the messages they deliver. One of Juliette's friends, Scottie, leaves her a message before he was killed:

"Putting more together. Don't trust porters, so wiring this. Screw costs, wire me back. Need transfr to Mech. Not safe here. – S." [59]

Thus, Juliette is not the only one to discover that the leaders intend to block any further exchange between people and that they kill covertly or openly in form of the cleaning death sentence everyone opposing to the system. All media, emails, Internet and even books, are censored as resources of information and as a means of communication. Elise, a surviving girl from the destroyed silo 17 collects articles of uncensored books of the past, an act which is strictly forbidden:

'What books were these?' the man in white asked. 'The ones with these animals, they were here in this silo?' […] 'Where are the books? It is so important, my daughter. There is only one book, you know. All these others are lies." [60].

In this restricted world, it is impossible to experience otherness as any information deviant from official statements. The inhabitants of the silos cannot reach independent information about their surroundings and their past, all they learn is the official statements and all they see of the surface is the camera's blurred vision of the dangerous outside world. The scenario in this trilogy is depicted in an extreme way, but shows the dangers of leaving human beings without independent information, communication and interpersonal exchange. It is simply cruel. Human beings are meant to discover and to interact, to express their own feelings and thoughts – these are very basic needs, that is way the discussion about Net Neutrality and the open Internet is such a heated debate.

3.2 "La possibilité d'une île" by Michel Houellebecq

Howey is not the only author to fear increasing restrictions on media and communication. Michel Houellebecq also imagines the survivors of a global catastrophe. The pseudo- scientific sect of the Élohimites invents cloned "neohumans" living isolated in their houses and when they die, they are replaced by their own clones. An online journal helps them to learn the memories of their previous "Self":

"Of two selfish and rational animals, the most selfish and rational of the two had ended up surviving, as was always the case among human beings. I then understood why the Supreme Sister insisted upon the study of the life story of our human predecessors. I understood the goal she was trying to reach: I understood, also, why this goal would never be reached." [61].

The quotation hyperbolically reflects Houellebeqc's aversion against the selfishness and egocentrism of mankind. This repeated memorizing instead of making own experiences leads to the loss of real emotions and the neohumans are reduced to rational thought and joyless immortality and to being trapped in their own egoism:

"On the other hand what they [the previous versions of Marie and Esther – editor's note] had known, and in a singularly painful way,

was nostalgia for desire, the wish to experience it again, to be irradiated like their distant ancestors with that force that seemed so powerful. […] Rejecting the incomplete paradigm of form, we aspire to rejoin the universe of countless potentialities." [62].

By only interacting socially via Internet and never meeting in "real life", the neohumans achieve, more or less voluntary, a state of peace of mind:

"According to the Supreme Sister, jealousy, desire, and the appetite for procreation share the same origin, which is the suffering of being. It is the suffering of being that makes us seek out the other, as a palliative; we must go beyond this stage to reach the state where the simple fact of being constitutes in itself a permanent occasion for joy; where intermediation is nothing more than a game, freely undertaken, and not constitutive of being. We must, in a word, reach the freedom of indifference, the condition for the possibility of perfect serenity." [63]

All exposure to otherness and all information depend in "La possibilité d'une île" on technology which easily can be controlled from the outside. Houellebecq's exaggerated vision of the future can be understood as warning against the growing importance and dependence on technologies which can easily be manipulated and censored. If the media are under the control of political, religious, economical or other dominant powers and if they censor and block the free exchange of information, the users would have no choice to access other points of view and therefore not be democratically responsible, when they communicate exclusively via technical devices. It might be one of the author's intention to sensitize for that problematic, or maybe he simply criticizes mankind in its disillusionment of being the rational and engineered species. For in trying to improve mankind with cloning, simplified nutrition and advanced media, the neohumans in his novel have lost important features which define being human: Emotions, empathy and social life – the dealing with the "other". For human beings do not exist on their own, but they need other human beings to interact, to meet the "other", an experience which creates external points of view on the own opinions and concepts. The Self becomes the object of its own reflections. Admitting the own objectivity helps to get a new perspective on oneself. Alterity therefore is needed to gain new points of view, to be able to relativize the own concepts and behavior.

In Houellebeqc's vision of isolated, mechanized clones, this outside perspective maybe is not possible. The neohumans never leave the safety of their houses and only continue their ancestor's life without really making new experiences and so is defined their social interaction. This is the extreme opposite of how media and Internet should be used as a tool to experience alterity. Instead, it is a recycling of old memories. This scenario reveals that media, as being used exclusively and replacing personal contact, might inhibit the experience of alterity instead of offering new perspectives.

3.3 "MaddAddam" trilogy by Margaret Atwood

The fact that the digital world is – still – openly accessible and manipulable is depicted as a threat in this case, in contrast to the novel "The Year of the Flood", the second book of the MaddAddam trilogy by Canadian author Margaret Atwood. In this trilogy, she portrays a pre-apocalyptic world reigned by ruthless corporations. Society as such seems intolerably degenerated and decadent, so one of the main characters of the first book ("Oryx and Crake"), the expert in biogenetics, Crake, develops a mortal virus to wipe out mankind and to restart evolution. Simultaneously, he creates a new transgenetic race of human beings who are immune to the virus.

Apparently, mankind has already lost its humanity, because it is ruled by excessive consumption, private corporations controlling every aspect of life. Instead of healing sickness, the corporations develop vitamin pills actually causing diseases to ensure that people are completely dependent of the pharmaceutical industry:

"She took more supplements, but despite that she became weak and confused and lost weight rapidly. […] All they did was poke at your tongue and give you a few germs and viruses you didn't already have, and send you home." [64]

The "normal" people of this dystopic world, the so-called "pleeb-rats", have no choice than believing in the corporation's promises, because they occupy any aspect of life, even the media. Thus, they have no access to independent information and communication. Any written word is a danger to its author, it is better to learn everything by heart. Transfer of information is therefore limited to mouth to mouth, which requires close relationships of trust and personal acquaintance. The corporations hinder the people in actually forming their own, independent and democratic opinions. The example of the medical system shows, how corporations decide over life and death. Other threats are the generally low inhibition threshold, with Gladiator fight, "Painball" [65], people being killed for organs and dead bodies being recycled into hamburgers [66], and of course the daily criminality in the streets. Hence, expressing own opinions, wanting to be different than the mainstream, leaving messages behind, can be rather dangerous:

"Beware of words. Be careful what you write. Leave no trails. This is what the Gardeners [a millenarian, Christian sect – editor's note] taught us, when I was a child among them. They told us to depend on memory, because nothing written down could be relied on. The Spirit travels from mouth to mouth, not from thing to thing: books could be burnt, paper crumble away, computers could be destroyed. […] As for writing, it was dangerous, said the Adams and the Eves [the sect's disciples – editor's note], because your enemies could trace you through it, and hunt you down, and use your words to condemn you." [67].

Best disguise is to assimilate to this environment. It is for this reason that the Gardeners avoid to use common communication tools and why young Amanda has developed a new art form of eaten messages which fade away:

"She'd [Amanda] written her name in syrup on the slab, and a stream of ants was feeding on the letters […] "It's neat," said Amanda. "You write things, then they eat eat your writing. So you appear, then you disappear. That way no one can find you." [68]

How are real experiences of alterity possible in this society? Mouth to mouth communication and memorizing are limited ways of communication requiring direct contact with the interlocutors. With the described threats and everyday crimes, one would, however, avoid the contact to someone of different opinions to not

risk one's own life, wouldn't he? Also, new ideas can only be spread by direct contact, that is why the Gardeners demonstrate in the streets while exposing themselves to the mob. The media are controlled by the private corporations as the dark and vague sensation of omnipresent evil indicates in the quotations above. Online resources, for instance, are dangerous to use, as shows the circumstance that the terrorist group MaddAddam, who believes to be the more effective – which means violent – opposition to the corporations, uses online games as cover for secret messages. Thus, opposite opinions to the mainstream can only be accessed by insiders and close friends. The experience of alterity in form of independent media and free exchange of opinions is not open to the public in this society reigned by the empty promises of seductive and ruthless corporations.

4. CONCLUSION

Summarizing the above sections, it can be said that in Hugh Howey's scenario, it is the government which controls every access to information and all communication to avoid alternative opinions and concepts. People are deprived of their human right to have independent exchange of information and media, it seems for the Greater Good of saving mankind, but the silo inhabitants are in fact reduced to guinea pigs of the leader's plan to sort out, in a life and death experiment, which silo is the healthiest to repopulate Earth. To prevent the people from uprising, they have no contact with alterity, with opposite concepts to the official version. Houellebecq criticizes how religion and science define the people's points of view. In his vision of the future, human beings are lonely, emotionless, communicating only online with each other, so they depend completely on the information passing through the technical devices. They never really meet another person outside of the digital world, which is an exaggeration of how we live today. Technology, instead of uniting people, isolates them. Atwood has a more positive view on technology and defends the freedoms of information and communication. In her dystopia, private corporations control everything, a horror vision which is also often depicted in the real debate about Net Neutrality. Losing free information and communication, having no access to alternative perspectives and to alterity as such seems, are, in conclusion of this paper, justified fears in our modern times. This globally perceptible fear even reflects on contemporary fiction. The continued back and forth on the political side, however, does not reassure the citizens.

5. ACKNOWLEDGMENTS

I would like to express my thanks to my supervisors Professor Rotraud von Kulessa and Professor Till Kuhnle, whose open minded understanding of research and unconventional ideas always inspire my own projects. Special thanks also to Andi, gifted programmer and most importantly, a good friend. Last, but not least I thank my husband and family for their love and support.

6. REFERENCES

[1] Kermode, F. 1966. *The Sense of an Ending. Studies in the Theory of Fiction*. Oxford University Press. Oxford. 44.

[2] Anders, G. 1959. *Über Verantwortung heute.* In *Die atomare Drohung. Radikale Überlegungen zum atomaren Zeitalter.* G. Anders. 2003[7]. C. H. Beck. Munich. 96.

[3] Krempl, S. EU-Abgeordnete knicken bei Netzneutralität ein. (June 2015). DOI: http://www.heise.de/netze/meldung/EU-Abgeordnete-knicken-bei-Netzneutralitaet-ein-2722819.html

[4] Kleinz, T., Beuth, P. *Es ist zu früh für einen Nachruf auf das Internet.* (July 2015). DOI: http://www.zeit.de/digital/internet/2015-06/netzneutralitaet-europa-trilog-spezialdienste.

[5] King, R. L. *FCC met with Canadian researcher to understand CRTC.* (July 2015). DOI: http://www.thestar.com/business/2015/02/26/fcc-met-with-canadian-researcher-to-understand-crtc.html.

[6] Cf.: ibid.

[7] Cf.: ibid.

[8] Cf.: ibid.

[9] (July 2015). DOI: http://www.lemonde.fr/neutralite-du-net/.

[10] (January 2015). DOI: https://www.google.de/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=neutralit%C3%A9+du+net.

[11] Ronfaut, L. *L'Europe moins ouverte que les États-Unis sur la neutralité du Net.* (July 2015). DOI: http://recherche.lefigaro.fr/recherche/access/lefigaro_fr.php?.archive=BszTm8dCk78atGCYonbyzmapnOM7xw740XXhvK4r0k6WtD7cfAEhCN594WmcQIQQu2IGtjAq08M%3D.

[12] Honoré, R. *Les États européens se déchirent sur la fin des frais d'itinérance pour le mobile.* (January 2015). DOI: http://www.lesechos.fr/tech-medias/hightech/0203973232895-les-etats-europeens-se-dechirent-sur-la-fin-des-frais-ditinerance-pour-le-mobile-1069012.php. The minister of Digital Affairs Axelle Lemaire: "Tout doit être fait pour affirmer le principe de neutralité du Net, tout en laissant de l'espace à l'innovation comme l'e-santé ou la TV très haute définition" "Everything must be done to affirm Net Neutrality, and at the same time letting space to innovation such as e-health and very high definition TV."

[13] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 10. (Heft Nr. 96).

[14] Cf.: ibid. 7.

[15] Cf.: ibid. 7.

[16] Cf.: ibid. 15. Refering to the study led by Scott Cleland *A First-Ever Research Study, estimating Google's U.S. Consumer Interest Usage & Cost* (2007-2010).

[17] Cf.: ibid. 12.

[18] Cf.: ibid. 28.

[19] Stiegler, Z., Spurmont, D. 2013. *Framing the Net Neutrality Debate.* In *Net Neutrality and the Fate of the Open Internet. Regulating the Web* Z. Stiegler et al. Lexington Books. Lanham. 132.

[20] Cf.: ibid. 131.

[21] Cf.: ibid. 136.

[22] Bortnikov, V. (2013). *Netzneutralität und Bedingungen kommunikativer Selbstbestimmung. Pflichten des*

freiheitlichen Verfassungsstaates zur Gewährleistung der Neutralität des Internets im Lichte der grundrechtlichen Schutzpflichtenlehre. C. H. Beck. Munich. 14-16.

[23] (July 2015). DOI: http://www.un.org/en/documents/udhr/index.shtml#a12.

[24] Bortnikov, V. (2013). *Netzneutralität und Bedingungen kommunikativer Selbstbestimmung. Pflichten des freiheitlichen Verfassungsstaates zur Gewährleistung der Neutralität des Internets im Lichte der grundrechtlichen Schutzpflichtenlehre*. C. H. Beck. Munich. 9.

[25] (July 2015). DOI: http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000020735432&categorieLien=id.

[26] (July 2015). DOI: http://www.un.org/en/documents/udhr/index.shtml#a19.

[27] (July 2015). DOI: http://www.cbc.ca/news/canada/telus-cuts-subscriber-access-to-pro-union-website-1.531166: „The site calls the company's move censorship, and TWU president Bruce Bell questioned its legality.“.

[28] Kimball, D. 2013. *When we talk about Net Neutrality: A Historical Genealogy of the Discourse of Net Neutrality* In *Net Neutrality and the Fate of the Open Internet. Regulating the Web* Z. Stiegler et al. Lexington Books. Lanham. 40.

[29] McCullagh, D. *Telco agrees to stop blocking VoIP calls.* (July 2015). DOI: http://news.cnet.com/Telco-agrees-to-stop-blocking-VoIP-calls/2100-7352_3-5598633.html.

[30] (July 2015). DOI: http://www.pcwelt.de/news/Schon-fuer-3G-Telefonate-genutzt-T-Mobile-blockt-Skype-auf-dem-iPhone-305280.html.

[31] Kremp, M. *Skype auf dem iPhone: T-Mobile blockiert Billigtelefonate.* (July 2015). DOI: http://www.spiegel.de/netzwelt/mobil/skype-auf-dem-iphone-t-mobile-blockiert-billigtelefonate-a-616587.html.

[32] Cf.: ibid.

[33] Bortnikov, V. (2013). *Netzneutralität und Bedingungen kommunikativer Selbstbestimmung. Pflichten des freiheitlichen Verfassungsstaates zur Gewährleistung der Neutralität des Internets im Lichte der grundrechtlichen Schutzpflichtenlehre*. C. H. Beck. Munich. 6.

[34] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 43. (Heft Nr. 96).

[35] Hoffmann, C. et al. (2015). *Die digitale Dimension der Grundrechte. Das Grundgesetz im digitalen Zeitalter.* Nomos. Baden-Baden. 124.

[36] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 45. (Heft Nr. 96).

[37] Hoffmann, C. et al. (2015). *Die digitale Dimension der Grundrechte. Das Grundgesetz im digitalen Zeitalter.* Nomos. Baden-Baden. 124.

[38] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 41. (Heft Nr. 96).

[39] Anderson, N. *Pearl Jam censored by AT&T, calls for a neutral 'Net*. (July 2015). DOI: http://arstechnica.com/uncategorized/2007/08/pearl-jam-censored-by-att-calls-for-a-neutral-net/.

[40] Marin, J. *La neutralité du Net à nouveau attaquée en justice aux Etats-Unis*. (July 2015). DOI: http://siliconvalley.blog.lemonde.fr/2015/03/24/la-neutralite-du-net-a-nouveau-attaquee-en-justice-aux-etats-unis/.

[41] Stiegler, Z., Spurmont, D. 2013. *Framing the Net Neutrality Debate.* In *Net Neutrality and the Fate of the Open Internet. Regulating the Web* Z. Stiegler et al. Lexington Books. Lanham.137.

[42] Kimball, D. 2013. *When we talk about Net Neutrality: A Historical Genealogy of the Discourse of Net Neutrality* In *Net Neutrality and the Fate of the Open Internet. Regulating the Web* Z. Stiegler et al. Lexington Books. Lanham. 35.

[43] Cf.: ibid. 35.

[44] Granjon, F. 2012. *Mobilisations informationnelles et Web participatif.* In *Internet et politique* A. Coutant. CNRS Éditions. Paris. 84-85. French original: “Les mobilisations informationnelles en ligne se présentent comme des sites de prise de parole travaillant à la réappropriation du débat public. […] Aussi, Internet a-t-il souvent été présenté comme le dispositif technique de la parole et de la diffusion étendue des informations au plus grand nombre, hors des zones usuelles d'exercice de la citoyenneté […] [les blogs] sont précisément ces formes de production d'information citoyenne à l'initiative d'individus non affliliés (à des partis ou des associations traditionnelles)”.

[45] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 7. (Heft Nr. 96).

[46] Cf.: ibid. 7.

[47] Doctorow, C. *If one thing gives me hope for the future, it's because of internet freedom.* (July 2015). DOI: http://www.theguardian.com/technology/2015/may/26/hope-future-internet-activism-freedom.

[48] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 24. (Heft Nr. 96).

[49] Cf.: ibid. 27.

[50] Martini, M. 2011. *Wie viel Gleichheit braucht das Internet? Netzneutralität zwischen kommunikativer Chancengleichheit und Infrastruktureffizienz.* Inaugural lecture. Speyerer Vorträge. 28. (Heft Nr. 96): “Das Internet eröffnet eine von institutioneller Einbindung und Zahlungskraft unabhängige Teilhabe an einem kollektiven Gestaltungs-, Kommunikations- und Willensbildungsprozess […] Als dezentrale Kommunikationsplattform nimmt sein kreatives Chaos eine immer bedeutsamer werdende Katalysatorfunktion in der Demokrasierung von Wissenszugängen ein.”

[51] Constantopoulou, C. C. 2013 *Agoras Virtuelles : la « démocratie » contemporaine* In *Les réseaux sociaux sur*

Internet à l'heure des transitions démocratiques S. Najar. IRMC. Tunis. 465-66.

[52] Cf.: ibid. 471.

[53] (July 2015) https://www.youtube.com/watch?v=A1zLzWtULig.

[54] Vanbremeersch, N. (2009). *De la démocratie numérique*. Seuil/Presses de Science Po. Paris. 51-53: "C'est un espace alternatif, identitaire, et un lieu d'entreaide […] en ligne, c'est essentiellement la formation du jugement. C'est là que se crée l'opinion, pour toute foule de lecteurs et de blogueurs. On lit et on discute pour se frotter à autrui et comprendre."

[55] Engélibert, J-P. (2013). *Apocalypses sans royaume. Politique des fictions de la fin du monde, Xxe-XXIe siècle.* Classiques Garnier. Paris. 10.

[56] Howey, H. (2013). *Wool* Arrow Books. London. 28-29.

[57] Cf.: ibid. 30-31.

[58] Cf.: ibid. 182-183.

[59] Cf.: ibid. 191.

[60] Howey, H. (2013). *Dust* Arrow Books. London. 302-304.

[61] Houellebecq, M. (2007). *The Possibility of an Island.* Vintage Books. New York. 336. French original: Houellebecq, M. (2013) *La possibilité d'une île*. J'ai lu. Paris. 446: "De deux animaux egoïstes et rationnels, le plus egoïste et le plus rationnel des deux avait finalement survécu, comme cela se produisit toujours chez les êtres humains. Je compris, alors, pourquoi la Sœur suprême [a dubious religious leader of the Élehomites -editor's note] insistait sur l'étude du récit de vie de nos prédécesseurs humains ; je compris le but qu'elle cherchait à atteindre. Je compris, aussi, pourquoi ce but ne serait jamais atteint."

[62] Cf.: ibid. 294-295. French original: Houellebecq, M. (2013) *La possibilité d'une île*. J'ai lu. Paris. 392-393: "Ce qu'elles [the previous versions of Marie and Esther – editor's note] avaient par contre connu, et cela de manière singulièrement douloureuse, c'était la nostalgie du désir, l'envie de l'éprouver à nouveau, d'être irradiées comme leurs lointaines ancêtres par cette force qui paraissait si puissante. […] Rejetant le paradigme incomplet de la forme, nous aspirons à rejoindre l'univers des potentialités innombrable".

[63] Cf.: ibid. 260. French original: Houellbecq, M. (2013) *La possibilité d'une île*. J'ai lu. Paris. 347: "Selon la Sœur suprême, la jalousie, le désir et l'appétit de procréation ont la même origine, qui est la souffrance d'être. C'est la souffrance d'être qui nous fait rechercher l'autre, comme un palliatif ; nous devons dépasser ce stade afin d'atteindre l'état où le simple fait d'être constitue par lui-même une occasion permanente de joie ; où l'intermédiation n'est plus qu'un jeu, librement poursuivi, non constitutif d'être. Nous devons atteindre en un mot à la liberté d'indifférence, condition de possibilité de la sérénité parfaite."

[64] Atwood, M. (2010). *The Year of the Flood*. Anchor Books. New York. 32-33.

[65] Cf.: ibid. 130.

[66] Cf.: ibid. 43: "They also ran corpse disposals, harvesting organs for transplant, then running the gutted carcasses through the SecretBurgers grinders".

[67] Cf. ibid. 7.

[68] Cf. ibid. 100.

The ethics of human-chicken relationships in video games: the origins of the digital chicken

B. Tyr Fothergill
School of Archaeology and Ancient History
University of Leicester, Leicester
LE1 7RH, United Kingdom
+44 0116 223 1014
bf63@le.ac.uk

Catherine Flick
De Montfort University
The Gateway
Leicester, United Kingdom
LE1 9BH, United Kingdom
+44 116 207 8487
cflick@dmu.ac.uk

ABSTRACT
In this paper, we look at the historical place that chickens have held in media depictions and as entertainment, analyse several types of representations of chickens in video games, and draw out reflections on society in the light of these representations. We also look at real-life, modern historical, and archaeological evidence of chicken treatment and the evolution of social attitudes with regard to animal rights, and deconstruct the depiction of chickens in video games in this light.

Categories and Subject Descriptors
K.4.0 [**Computers and Soceity**]: General

K.4.1 [**Computers and Society**]: Public Policy Issues: Ethics

K.8.0 [**General**]: Games

General Terms
Human Factors, Theory

Keywords
Chickens, Video Games, Archaeology, Human-Animal Interactions

1. INTRODUCTION
The chicken (*Gallus gallus domesticus*) is the world's most abundant bird; it is symbolic of both domesticity and high-tech food production. Globally, we consume millions of tonnes of chicken flesh and eggs; we also keep them as pets and they play roles in spiritual practices. It is no wonder, then, that the chicken features in our technological representations of fantasy worlds in video games. In fact, the chicken is a common figure in a wide range of video games, where it is chased, killed, kicked, choked, ridiculed, used as a comedy prop, eaten, and required to endure various other indignities. In some games it represents a more positive symbol of abundance and wealth, including racing, breeding, farming, and riding; chickens exact revenge on overly bloodthirsty chicken killers. The obsession with chickens (or creatures with chickenlike qualities) in video games is representative of a longer tradition of representations of chickens in media.

The complexity and contrasts of the digital chicken reflect the similarly multi-faceted past of the species, perceived as a domestic animal with many useful purposes and an imagined, depicted being. In this paper, we explore the many and varied roles and uses of the chicken in video games and contextualize these with archaeological and historical data.

2. THE DOMESTICATION AND SPREAD OF *Gallus gallus*, THE CHICKEN
Humans have conceptually and physically shaped and re-shaped the other animal species with which we have interacted; few examples of this are more striking than the chicken. Domestication is often conceived of as an activity undertaken by humans which converts a wild plant or animal into something else, a living thing entirely under the control of or dependent upon humans to survive. The complexities of such a transformation are immense, and are more accurately framed as "an ongoing co-evolutionary process rather than an event or invention" [15].

The primary wild progenitor of the domestic chicken is the red junglefowl (probably with some genetic input from the grey junglefowl), which was domesticated by the 6th millennium BC somewhere in Southeast Asia; there were probably multiple centres of domestication [41]. From the outset, there have been multiple forms of chicken-human interactions; this is well-reflected in current relationships between people and chickens as well as in video games.

3. ROLES OF THE CHICKEN IN PAST HUMAN LIFE
Although chickens are frequently thought of as a meat source today, the chicken may not have been initially domesticated for food. It may be that the bird was kept for other reasons, including cockfighting [44]. The bounteous gifts of the chicken to past humanity are not limited to flesh and "fun", however; chickens were sacrificed as parts of sacred and divinatory practice [18]; various parts of the chicken have been utilised for medicinal purposes [23]; hen's eggs have long been a source of sustenance and characterised as "the world's most versatile" culinary ingredient [34]; and chicken feathers have been used for bedding [28]. Chickens were entangled with ancient divinity and the afterlife, subjected to violence, commodified, associated with both women's work [3, 35] and men's play [42, 44], and linked to both bravery and cowardice. These strands are complex, interwoven and often contradictory; even in considering them below, we outline but a fraction of the ways in which chickens played a role in the past human experience.

3.1 The Sacred Chicken

Images of cocks, sometimes paired as if to fight and often with exaggerated features, are prominent in ancient funerary décor. Roman gravestones feature these depictions in numerous locations and solitary male chickens top the stepped tombs of Carthaginian North Africa [13]. Across diverse parts of the globe including south-east Asia, Africa, and Europe, chickens were buried with humans in ways which do not suggest placement as "food offerings" [17, 40, 41]; [Sykes pers. comm.]. Chickens also played a vital role in terms of connections to the divine. At a centres for the Asiatic cult of Mithras, thousands of chicken remains were excavated, the majority of these from cocks [22]. Roman deities such as Zeus and Mercury also had links to the species, and portrayals of these gods often include cocks. Livy's record of omens includes cocks changing into hens and hens changing into cocks [24]; this hints at how different male and female chickens are in appearance, a point to which we will return with regard to portrayals in video games. The last words of Socrates purportedly included a request for the sacrifice of a white cock to Aesculapius, the god of medicine, to whom offerings were typically given for recovery from ailments [46]. *Kapparot*, the sacrifice of a cock on the eve of Yom Kippur, the holiest day in the Jewish calendar, was intended to transfer the sins of the individual to the chicken as part of a Day of Atonement (*Shulchan Aruch Rama O.C.* 605:1). Ghanaian religious practices in the Tongo Hills still require the sacrifice of chickens, often of a specified colour, at certain shrines or in household contexts [18]. The contents of a Talensi diviner's bag include the foot of a chicken [18]. In Christianity, the crow of the cock was a critical temporal measure of Peter's denials of Jesus, and it would herald the return of Christ. Christ uses the idea of a hen gathering her chicks to illustrate his feelings about the people of Jerusalem (Matthew 23:27; Luke 13:34) and such a scene is the subject of a mosaic on the altar of the Church of Dominus Flevit in Jerusalem. A papal edict issued by Nicholas I in the 9th century required all churches to use only the image of a cock as their steeple weathervanes. A popular emblem of Portugal is O Galo de Barcelos, an icon derived from a legendary cock which is said to have crowed (despite having been roasted!) to prevent the hanging of an innocent man who was on pilgrimage to Santiago de Compostela.

3.2 The Violence Inherent in the System

Direct archaeological evidence for violence involving chickens is rare, and palaeopathology (the study of past disease and injury) may be of little help as the chickens involved in such activities stood little chance of surviving their injuries. Historical sources clarify the details of two specific practices: cockfighting and cock-throwing.

Themistocles, an Athenian general (*c.* 524-460 BC), is often credited with popularising cockfighting amongst the Greeks and the Western world thereafter, but it probably originated in Southeast Asia and was perpetuated by various groups along the way. The concept of fighting cocks was used by the Greeks in theatrical scenes painted on red-figured Attic ceramics dating to the first half of the 5th century BC. In one illustrated scene from Aristophanes's comedy *Birds* (J. Paul Getty Museum, Malibu 82.AE.83, Side A), young actors are shown to be dressed as male chickens and engaged in some form of combat or argument [10], an early hint at the chicken costume's association with humour. Fighting cocks are also a frequent subject of Roman art and mosaics, and the activity of cockfighting retained an association with men, often members of the military, nobility, and occasionally royalty. Henry VIII had a cockpit installed at Whitehall, which was (after a fire in 1697), converted to the Privy Council room [42]. As the popularity of cockfighting increased and people at most levels of society began to take part, the association with the upper class waned and it was made illegal in the entirety of the United Kingdom in 1895. Cockfighting continues to be a popular activity in a range of communities across the globe, and remains legal in several countries.

The violent practice of "cock-throwing" was popular in England from at least 1409 [38]. Throwing at cocks, also called "cock-threshing", "cock-stele", and "cock-running" was associated with Shrovetide, an opportunity to engage in blood sport and other diversions in the three days before Lent [38]. Cock-throwing involved restraining a male chicken in some way, (e.g. tying it to a stake or placing it in a ceramic vessel) and then pelting it with sticks, stones and other objects until the creature died. If the cock's leg broke as a result of a particularly brutal beating, it was often propped up so that the "game" could continue to completion [42]. It was a deeply popular pastime, as suggested by the revolt of apprentices in Bristol in 1660 when the local Quaker officers forbade them to engage in it [43]; people continued to "throw at cocks" in England until at least the 18th century [37].

The acceptability of cock-throwing began to decline toward the end of the early modern period. The English painter William Hogarth presented it as the first stage of cruelty his series of paintings *The Four Stages of Cruelty* (1751). Modern symbols such as le Coq Sportif and the logo for Tottenham Hotspur are taken directly from the culture of cockfighting and are also associated with masculinity and male activities. Of the many different roles of the chicken in the past, these acts of abuse (by current welfare standards) are clearly reflected in various digital worlds. We discuss this further in sections 5.2 and 6.2 below.

3.3 Chickens as Product

Chicken bones from many archaeological sites show evidence of culinary processing, including butchery marks and burning, but even as meat or an efficient source of protein through eggs, the chicken represented more than simple sustenance. The chicken was the essential economic unit underpinning the ancient economy of Kellis in the Dahkleh oases, and a specific group of husbandry specialists were dedicated to their upkeep [13]. There is archaeological evidence for large-scale poultry production elsewhere in Egypt including artificial egg-hatching technology [12] and cooperative networks of poultry farms [13].

3.4 Chickens and Domesticity

Historically, chickens as a species are firmly associated with the domestic sphere, the household, ideas of safety and a welcoming, secure environment. These constructs are clearly linked to the "feminine ideal" and ideas about "women's work" [3, 35]. Yet, the group responsible for the husbandry of the chickens at Kellis included men; furthermore, the individuals taking part in poultry production networks in ancient Egypt were also men [13]. In short, any assumption about the role of gender in animal husbandry practices in the past should be interrogated.

3.5 Chickens and Cowardice; Chickens and Bravery

Like the idea of chickens as solely the domain of women, "chicken as coward" has not always been a universally-accepted concept. There are current examples of this: someone who won't go through with something is said to "chicken out"; to "be chicken" or "chicken shit" is to be afraid, etc. From at least the

mid-19th to the mid-20th century, "chicken-hearted" referred to a wretched, craven or cowardly individual [6, 26]. The French phrase "poule mouillée" ("wet hen") translates to wuss or weakling, whilst "wet hen" in British English is used to mean someone sad, useless and a bit of a "wet blanket" [32].

In contrast, terms and phrases related to bravery, aggression and success such as "cock-sure", "cock of the walk", "cocky", "to rule the roost, "to play chicken" (faceoff), and "live like fighting cocks" (to feast well, [4]) often relate to male chickens and are associated with masculine activities such as cockfighting.

Chickens were the subject of comedic focus and jokes as well as zoomorphic projection in the past, though little in the way of material evidence exists to prove this practice. Their appearance in humorous literature is early: in the Greek comedy *Birds* by Aristophanes, first performed in 414 BC, brave or strutting cocks were used as illustrative devices for behaviour (lines 1105-1109) and entire empires (lines 616-624)[2]. The earliest modern chicken joke was published in a New York magazine called *The Knickerbocker* [47] the progenitor of the familiar "Why did the chicken cross the road?".

3.6 Cycles of Perception and the Chicken

The way in which humans perceive other animals impacts how we treat and raise them, which in turn leads to physical changes in some animals (e.g. increased size, rapid growth), which then perpetuate and deepen our views of that species. These human perceptions are reflected in the digital world; chickens still entertain us in video games, even if we're not hurling sticks and stones at them. A lack of understanding of past chicken-human relationships can lead to portrayals in video games which may only serve to normalise negative ideas about the real animal.

It is from this that the current paper emerges: how are the ways in which chickens are portrayed in video games linked to perceptions of chickens in the past? What does our obsession with and representation of chickens in video games say about our society? Is it right to represent chickens (and other domestic animals) in video games as objects of brutalisation?

4. METHODOLOGY

This study takes a qualitative critical approach to answer the research questions in two stages: firstly by identifying video games with portrayals of chickens and "chicken-like" entities, and analysing the roles that these chickens play within the games, and secondly discussing the findings of the first aspect in the context of Internet forum discussion of chicken-related play in games. It uses the lenses of the five categories identified above: sacred chickens, violence and chickens, chickens as product, chickens and domesticity, and linguistic conceptual/human behaviour references to chickens.

53 video games were identified as having chicken-related aspects. These games spanned the video game timelines from the early 1980s through to recently published games (2015). Although this is not an exhaustive search for chicken representations in video games, we believe that the data collected sufficiently represents the categories we determined above. A full list of video games can be found in the Appendix.

Searches on popular video game forums were made using Google to determine any references to chickens in any video game. These are critically discussed in the light of the categories and issues raised by the initial stage of the research.

This data collection and analysis approach allowed us to critically reflect on the research questions and focus on in-depth data analysis rather than a shallower and unrepresentative quantitative approach.

The critical approach used is an ethical analysis of the representation of chickens in video games, how these are linked to historical and modern societal perceptions, uses and abuses of chickens, and discussing whether such representations are ethically acceptable in video games.

5. ANALYSIS: CHICKENS IN VIDEO GAMES

This section describes all the roles that chickens play in video games, according to the categories identified in section 3. This will then feed into section 6 which will critically discuss these roles in the context of ancient and modern societal relationships with chickens.

5.1 Chickens as Sacred, Symbolic, Divine, and Magical Beings

In many video games, there are chickens with magical, divine or supernatural qualities about them – whether it's representations of chickens as gods, such as the god Egg-Tor in the *Fable* series, harbingers of doom or death, such as El Pollo Diablo in *Monkey Island* or El Pollo Grande in *World of Warcraft*, or possessors of other supernatural powers, such as being super-powered (*Mort the Chicken*; *Billy Hatcher and the Giant Egg*; *Far Cry 4*), magic-wielding (*Sly 3*; *Gauntlet Legends*), or undead (*Guild Wars 2*).

Figure : Chicken as religious iconography, *Forge Quest*

In some games, chickens are associated with more passive magical or symbolic functions, such as in *Guacamelee!*, wisdom – where a large chicken offers hints and gameplay suggestions. In *Forge Quest*, the chicken is revered in religious iconography (Figure). Eggs, particularly, are considered to have healing properties (it is assumed that the eggs in these games are chicken eggs – many have references to chickens and some have explicit connections between the chickens and eggs), such as in the *Resident Evil* series, but beware eating chicks in the *Fable* series, where in *Fable II* eating "Crunchy Chicks" can summon an evil temple or weapon, and in the third of the series eating chicks will make your avatar put on in-game weight and decrease its moral standing.

The eating of chickens and eggs will be covered further in section 5.3 which describes chickens as product.

5.2 Chickens and Violence

One of the most prevalent ways that video game players interact with chickens is through violence. Sometimes the chicken will attack the player (*Zelda* franchise, *Chuckie Egg*, *Resident Evil 5*,

Monsiegneur Cockburn, or in the case of the exploding hen mod for *Skyrim*, blow the player up), usually after the player has acted violently toward it. Some games allow fighting chickens to be trained and kept as pets to send into battle (*Final Fantasy* series, *Legend of Dungeon, World of Warcraft*, *Pokémon*). Some games have the player playing as the violent chicken, such as in the "ninja death chicken" and "macho chicken" mods for *Skyrim*, the gun-toting chicken mod in *Grand Theft Auto V* or protagonist Mort in *Mort the Chicken* and *Monsiegneur Cockburn* in the titular game. However, violence directed toward chickens is the primary defining relationship between chickens and the player.

Figure 1: Chicken kicking in *Fable Anniversary*

In many games, chickens exist as an object that can be injured or killed. In the *Fable* series, a "chicken kicking" competition allows the player to win awards, including a chicken costume (see Figure 1). This is similar to the chicken punching minigame in *Guacamelee!* where the player must punch the chickens into the correct bins to progress in the puzzle. In *Besiege*, the player creates machines that, as one of their aspects, can crush objects. A notable object that is crushable is the chicken, with an over-the-top blood spatter accompaniment. In *Far Cry 4, Grand Theft Auto V*, *Crysis* and *Counterstrike: Global Offensive*, chickens exist to be shot at and killed, with no particular reward. As mentioned previously, many games will also allow players to attack or threaten the chickens in a similar way, but the chickens will fight back.

5.3 Chickens as Product

Chickens are regularly portrayed as a "food" item, or lay eggs that can be used as food in video games, often with the result of regaining health (*Minecraft*, *Resident Evil* series, *Puzzle Craft*, *Chuckie Egg*, *Mort the Chicken, Harvest Moon*, *Farming Simulator*, *Castlevania*, *Monkey Island*, *Tekken 3*). In *Guild Wars 2*, there is even a special quest-related chicken that can spawn, called "Dinner". The entire roast chickens, located attached to the walls in *Castlevania* have become something of an in-joke in games with send-ups including *Dust: An Elysian Tail*, where health can be regained from a "mysterious wall chicken" (Figure 2). As mentioned previously, "Crunchy Chicks" can be eaten in *Fable 3* but will increase the avatar's weight and decrease its moral standing.

Figure 2: "Mysterious Wall Chicken" *Dust: An Elysian Tail*

Chicken (or in-game equivalent) feathers are also sometimes used in crafting in some "survival" games such as *Minecraft* and *DayZ Standalone* and MMOs such as *Final Fantasy XI/XIV*. They are often used for making fletchings for arrows, making or decorating armour, quills, and other similar uses.

5.4 Chickens and Domesticity

Chickens are often represented "realistically" in domestic settings in games, i.e. within villages or farms in appropriate housing (such as *DayZ Standalone*'s chicken coops, roaming around villages in *Crysis, Skyrim*, *Forge Quest*, *EverQuest 2* and *Far Cry 3*, or being caged in *Resident Evil* games). In *World of Warcraft* there is a quest where the player needs to make the chickens feel comfortable enough to lay an egg. In *Guild Wars 2* there are quests that increase your renown that involve dealing with domesticated chickens, such as returning chickens to a pen. Similar "round up" quests can be found in the *Zelda* series, *Vanguard: Saga of Heroes*, *Mort the Chicken*, *Fable*, and *Guacamelee!*.

Chickens can also be bred in many games, such as in *Final Fantasy* for riding and racing, and in *Harvest Moon*, *Puzzle Craft*, *Minecraft* and *Farming Simulator* to simulate real world farming practices. *Divinity: Original Sin* has realistic depictions of chickens in that sex ratios are relatively accurate and the graphics are true-to-life with respect to sexual dimorphism. In *Banjo-Tooie* a hen character Heggy lives in an Egg Shed and will hatch eggs brought to her by the player. Hen House Harry theoretically looks after egg production in *Chuckie Egg*. They are also kept as pets or companions, such as in *World of Warcraft*, *Skyrim*, and the *Final Fantasy* series, and you can find Egbert the chicken villager in *Animal Crossing*.

Chickens and chicken-like creatures can be ridden as mounts in *Final Fantasy* (Figure 3), *World of Warcraft*, and *Rift*.

Figure 3: Fat Chocobo mount in *Final Fantasy XIV*

Amusingly, in *Skyrim*, chickens can potentially be witnesses to the player's criminal behavior, so if a player wishes to eliminate all witnesses, they must also kill any chickens (normally seen as harmless unimportant creatures) who saw them in the act. If they do not do this, the guards in the town will stop the player and arrest them for the crime.

5.5 Chickens as Illustrations of Human Behaviour; Chickens as Jokes

In some games, the chicken is related to bravery, for example, *Billy Hatcher and the Giant Egg* where chickens are seen as courageous. The opening cinematic for *Fable III* follows a brave but ultimately doomed chicken which is battered and beaten through an industrial-revolution themed city attempting to gain its freedom. Chickens are (perhaps) brave but doomed initial experiment subjects in *Portal*. Unfortunately, the chicken is not as lucky in other games, where it, or aspects related to it, is seen as lazy and shiftless (*Animal Crossing*), cowardly and cheating (*Far Cry 3*, *Fable III*), or associated with questionable sexual tastes (*Witcher 2*).

Rubber chickens appear in games as well, with the most notable example being the rubber chicken with a pulley in the middle from the *Monkey Island* series. Originally thought to be completely useless, it transforms into a remarkably useful item. A rubber chicken mod is also available in *Skyrim*.

Other joking aspects of chickens included in games include an *Arrested Development* reference in *Rift* where players' avatars can use a /chicken emote to dance like a chicken, some variations of which are homages to the *Arrested Development* characters' humorously terrible portrayals of chickens in the show.

Chicken costumes also appear, sometimes giving bonuses to "silliness" or negatives to "attractiveness" (such as in *Fable III* and *Fable Anniversary*). In *Hitman 2: Blood Money* an elite group of assassins wear a chicken-like outfit. In *Witcher 2*, the player can win a chicken beak mask as a reward for fulfilling a rather bizarre storyline to do with a chicken fetishist. Chicken or similar costumes also abound in *Final Fantasy* MMOs as event awards.

In *Orcs Must Die 2*, a ring of polymorph can transform an enemy into a particularly harmless creature (in a joking way) – a chicken! The chicken represented in this game also has an oversized cloaca (exit orifice), presumably part of the joke.

Additionally, in *Far Cry 3*, the harmless, easily scared chicken has an entry in the handbook "survival guide": "Chicken is chicken, you'd have to be from some backwater like Canada to not know what chicken is. And chicken is un-American. Us true patriots eat only 100% U.S.A. Kobe beef." (The other joke presumably being that Kobe beef is Japanese.)

Of course, the 19th-century joke about how the chicken crossed the road was taken literally by *Freeway* as the player must navigate the chicken through a busy freeway crossing.

Finally, in what is probably mostly unintended to be a joke, most of the chickens referred to as male in these games are actually hens. For example, the lovingly-created *Skyrim* companion chicken mod, rendered in immaculate detail, is decidedly a hen, but referred to as "he" throughout the game. Also in *Skyrim*, the "macho chicken" mod, although referring to masculine attributes, has all hen heads. The decidedly traditionally masculine-acting gun-toting chicken mod in *Grand Theft Auto V* is also a hen.

6. DISCUSSION

6.1 Chickens and Masculinity

Video games are not a neutral form of entertainment. Traditional gender roles are often developed and reinforced in video games, and socialize young people in expectations for their gender; gendered play spaces are the new norm, with "boy culture" moving into virtual play spaces instead of remaining outdoors as in previous centuries [7]. Indeed, male play space is intensified in video games with the player's "physically active role in controlling the central protagonist" [20], many of whom are male characters with high levels of machismo. Masculinity is a complex concept, and video games tend to fall into the trap of portraying masculinity (and male characters) as part of the hegemonic masculinity of macho, domineering, rigidly "manly" men. In one study, male video game characters were found to be far more aggressively depicted than female characters [11], and depictions of such hyper-masculine traits can directly influence young men's beliefs in acceptability of such traits as ideally masculine [36]. This hegemonic masculinity is largely criticized in the literature, as it is not the reality of men and male behavior, and can in fact be detrimental to men [9]. Instead, Connell and Messerschmidt argue for a usage of "masculinity" to encompass more than just a set of toxic, rigid traits, and to look at contextual and positive depictions of masculinity. Here, we examine how depictions of chickens in video games and male associations with chickens can potentially contribute to the detrimental, hegemonic theory of masculinity.

Thus, incursions into the video game space by the joke of a chicken may be more serious than they initially seem. In the gun-toting chicken mod for *Grand Theft Auto V* the protagonist is changed into a hen. It can perform all of the actions that the usual human (male) protagonist can perform, such as stealing cars, shooting people, etc. It tucks the gun behind its wings and can be seen holding up shops and running people over in the video released by the mod's developer [16]. This is a humorous mod because it takes a usually benign and seemingly harmless animal and puts it into a heavily violent situation (running over pedestrians, shooting at police, etc.). However, it could be seen as

a natural extension of traditional cockfighting, cock-stele and other historical situations of violence and machismo that chickens have found themselves in (see more in section 6.2). In this game, the chicken finds itself in the stereotypically traditional male protagonist role; perhaps allowing the players of the game a humorous way to reclaim some of their perceived hegemonic masculinity by controlling a less macho version of the protagonist.

In *Mort the Chicken*, the hero's chicken-dominated world is invaded by sentient cubes which steal chicks. Mort the rooster has to use his super powers (and comb-whip) to reclaim the chicks. A "ruthless commando" (according to the leader of the cubes), Mort pecks eggs for power-ups and collects the chicks which flock around him as he flaps through the level. Once again, as in *Grand Theft Auto V*, the chicken takes the place of the human male protagonist, with Mort taking on the role of humorously macho non-man. Moreover, it is obvious that this is done in a joking fashion, perhaps with the same intent as *GTAV*.

A lack of understanding with regard to chicken sexual dimorphism (differing appearance between sexes) adds another layer of interpretive complexity. For example, the Skyrim chicken companion mod, despite being carefully crafted, provides a hen for what is intended to be a male chicken. Many games in which chickens are visual indicators of safety feature only hens. An exception to this is *Monsiegneur Cockburn*, a *Doom* clone evocative of a cockfight, in which the player controls a cock who kills other continuously-respawning cocks in a pit.

The "macho man" chicken mod for *Skyrim* [25], allowing the protagonist to play as a half-chicken, half-man avatar, with sounds replaced by a "Macho Man" Randy Savage voiceover, reaffirms Kirkland's concern for a masculinity that is stereotyped by muscular machismo. However, in some ways it is turned on its head (literally) through the graphic of a hen's head. Although chickens can exhibit bilateral gynandromorphy (the condition of having one half of the body biologically male and the other female), this parallel is probably unintended. The obviously amusing bent to this is the juxtaposition of a "macho" type image and sounds with a seemingly ridiculous animal head. This combination could either conjure up a masculinity-related link with cock-fighting and other violent chicken-related diversions, or an appeal to a more traditionally feminine domesticity. The fact that "Macho Man" Randy Savage, a famous wrestler, is the inspiration for the mod, makes the juxtaposition all the more bizarre. Indeed, comments on the mod's homepage indicate that this mod is considered to be quite disturbing: "Those preview pics [sic] are going to give me nightmares. Great yet disturbing mod"; "what is this i [sic] am scared"; "if it were not so creepy i [sic] would use it for real" [25]. Although this mod is clearly made as a joke, it retains some stereotypically hegemonic masculine approaches that reinforce traditional, rigid male roles (obvious musculature, wrestler voiceover).

Figure 4: "Macho Man Chicken" mod, *Skyrim* [25]

The masculinities depicted in video games tend to mirror those of traditional male-chicken relationships, however jokingly – rigid traits of machismo, aggression, and dominance over the environment and negative reinforcement of the desirability of such traits in male players. The humour of the scenarios adds to the negative reinforcement by drawing on the more modern understanding of "being chicken" and the association of chickens with cowardice as discussed in section 3.5.

These concepts are extremely well summarized by the upcoming *Metal Gear Solid V* in which there is an item, a "chicken hat" which grizzled veteran Solid Snake can wear if the player is finding a particular mission too difficult – with the hat on, enemies see Snake as a chicken (an insignificant object to ignore); it is implied that the player is a coward for needing to wear the hat.

6.2 Chickens and Abuse

Violence against chickens is currently unacceptable by recent welfare standards. This expectation extends not only to special breeds of chicken or pets; even broiler chickens are expected to live free from overt violence. A spectacular public furore resulting from the release of a video documenting routine stomping and kicking of chickens at a US supplier for Kentucky Fried Chicken [19] is one example of the acceptability gap between chicken treatment in reality and in video games.

In video games where violent acts against chickens are presented as a fun competition or form of diversion, e.g. chicken kicking in the *Fable* series, it could be argued that the inherent welfare perceptions reflected therein more closely resemble those of 17th-

century Bristolian apprentices rioting for their Shrovetide cock-throwing than modern, presumably enlightened heroes.

Although the archaeological and historical evidence supports the reality of the chicken (especially the cock) as an aggressive, strutting, fighting powerhouse, players do not expect a chicken to "strike back", perhaps due to overriding preconceptions about the cowardly, unimportant, disposable nature of chickens. Chickens in games which are very powerful or respond in kind to violence (section 5.2) are therefore intended to be unexpected, which further entrenches the conceptualization of the chicken as a simple object which a player can attack "to see what happens" or because it is perceived as humorous. In response to a query on the presence of chickens in video games, forum user "Soghog" on videogamesawesome.com (2012) replied: "Because chickens are funny. Abusing chickens is funnier".

6.3 The Societal Importance of Chickens

The ubiquity of chickens in video games reflects the ubiquity of the chicken globally. The chicken is a vital source of nutrition on a global scale. In 2013 alone, 21.7 billion chickens were produced for human consumption and 68 million tonnes of chicken eggs (FAOSTAT).

Chickens have been bred for desirable traits (certain colourations, comb shapes, numbers of toes and feathered crests) for thousands of years. Archaeological chicken crania with a pathological condition called a cerebral hernia (present in some breeds presenting feathered crests) have been excavated from a range of sites dating to the Roman period onward in Europe [5, 14]. Chicken breeding became widespread in the 19th century and showing continues to be a popular pastime.

A project called Hen Power, designed to combat loneliness and isolation amongst the elderly, encourages pensioners to rear and care for chickens on a daily basis and has met with great success [8]. The inclusion of complete, individual chickens in human burials (discussed in section 3.1) suggests that chickens have been our companions in life and in death for thousands of years. Indeed, Honorius, the Western Roman Emperor (AD 393-423) had a pet chicken named Roma, whom he reportedly doted upon [33].

Chickens retain great symbolic importance with regard to a range of spiritual practices (section 3.1). It is difficult to detect whether deific, supernatural, and sacred manifestations of chickens in video games are somehow connected to this ancient association or are presented in these ways as an ironic joke by designers who sought out what they perceived to be an unremarkable creature.

All of the aspects (food, special breeds, companionship, spiritual practice) present in video games are reflected archaeologically. The cycles of perception mentioned in section 3.6 have shaped real chickens and video game chickens in turn; like a broiler, video game chickens are often short-lived and viewed as completely disposable. This is not always the case, and we present a number of instances in which chickens are heroic, wise and worthy of admiration (section 5.1).

6.4 Ethical Representations of Chickens?

In considering the more negative representations of chickens in games as discussed above, it is important now to discuss the ways that video game developers could depict chickens responsibly, in order to build a more ethical relationship between humans and chickens. While some of these ideas may not be particularly interesting as the subject of games, or might be considered a little far-fetched, the purpose of this section is to provide some foci of reflection for developers including chickens in their games.

Most of the farming simulator genre of games (e.g. *Farming Simulator*, *Harvest Moon*, *Puzzle Craft*) have chickens as a farmable item, and many games (as mentioned in section 3.3) include chicken as a food item. On the one hand, it is important to emphasise sustainable, ethical practices as a normative expectation for farming, such as free range eggs, ethically treated animals, quality feed, etc. However, on the other hand, there could be the opportunity to highlight issue with unethical farming, such as battery hen farms, factory farms, or high density barn farmed chickens [1, 39]. It could suggest more sustainable methods for farming, such as a reduction in meat consumption, or more stringent regulation of treatment of chickens in farms. This, in turn, could also improve understanding of the risks that lead to the emergence and transmission of chicken-human diseases such as A(H5N1) and A(H7N9) avian influenzas.

Another way that chickens could be portrayed more ethically is to remove them as objects of abuse. Certainly, it might be funny for players to be able to kick, shoot, grind, and mash chickens in games, but is it really necessary? In the real world, animal abuse is a complex subject with some groups (notably animal rights groups) claiming it is a predictive factor for future violence, and others claiming that it has no effect on future violence [21]. Also important to mention is that violent video games are not linked with real-world violence [27]. However, this does not mean that cruelty to animals should be normalized – or even glorified – within games. For many gamers, it can be quite a distasteful experience where killing or maiming animals is part of the game (see [31]). PETA has campaigned against glorifying animal cruelty in video games [29, 30]. Although some of these complaints may seem far-fetched, as digital animals are not "real", Hochschnarter rightly points out that uncritical portrayals of violence toward animals could lead to normalization of violence against animals, and video games' increasing market share in media means that they should be criticized [45].

Finally, disassociation of the chicken from negative aspects of masculinity in video games would be another general improvement. We have already explained the problems associated with this, and more positive, productive, and ethically acceptable representations of masculinity in video games would benefit men as well as women. In video games, this would mean critically assessing the use of "chicken as joke" aspects – particularly regarding masculine traits, and reassessing traditional macho roles for men in video games. Perhaps it is time to bring back "cock-sure" and "cocky" associations rather than "being chicken"? After all, as we saw in *Fable III*, and to a lesser extent in *Mort the Chicken*, as well as in the fighting companion/pet depictions, chickens can certainly be extremely brave in video games.

7. CONCLUSIONS

In this paper, we have shown how complex relationships between chickens and humans has been recreated and perpetuated in video games. We have shown how the representation of chickens in video games can reinforce a negative concept of masculinity, with depictions of machismo, aggression, and dominance over the environment, and with jokes about chickens in games adding a negative reinforcement of the desirability of such traits in male players. We have also provided evidence for the close resemblance of attitudes towards chickens in video games to historical attitudes now considered inhumane. There is a spectrum of human-chicken relationships which is well-represented in video games, but some aspects have been distorted or lost in translation,

e.g. a lack of understanding and accurate representation of chicken sex despite the fact that the birds are sexually dimorphic. In other cases, video games reflect outdated and cruel attitudes to chickens in situations which are not far removed from cock-throwing, and certainly are not in line with modern views on chicken welfare.

As we have seen, chickens *can* be represented ethically in video games – either through a holistic approach that depicts them in a reflective context (such as the vengeful chickens in *Zelda* that show surprising realism in their ferociousness, the simulated farming in *Minecraft* or *Farming Simulator*), or through explicitly *not* depicting them as objects of unnecessary violence (for example in *Divinity: Original Sin* where killing chickens upsets villagers around you) and thus not contributing to a normalization of cruelty to animals.

This paper contributes to the understanding and analysis of video games by looking at them from a holistic perspective incorporating historical and archaeological understandings of chickens, and discussing the relationships these representations of chickens have with a modern, video-game-playing society.

8. ACKNOWLEDGMENTS

This work was supported in part by the AHRC-funded project Cultural and Scientific Perceptions of Human-Chicken Interactions.

9. REFERENCES

[1] Anomaly, J. 2014. What's Wrong With Factory Farming? *Public Health Ethics*. (Feb. 2014). DOI=10.1093/phe/phu001

[2] Aristophanes 2008. *Birds*.

[3] Bourke, J. 1993. *Husbandry to housewifery: women, economic change, and housework in Ireland, 1890-1914*. Clarendon Press ; Oxford University Press.

[4] Brewer, E.C. 1898. *Dictionary of phrase and fable*. Henry Altemus Company.

[5] Brothwell, D. 1979. Roman evidence of a crested form of domestic fowl, as indicated by a skull showing associated cerebral hernia. *Journal of Archaeological Science*. 6, 3 (1979), 291–293.

[6] Broughton of Berryhlonsworth, W.H. *Letter from W.H. Broughton of Berryhlonsworth to Hamilton Hume, Esq*. Public Record Office of Northern Ireland. Item Reference Number D2765/D/29.

[7] Cassell, J. and Jenkins, H. 2000. *From Barbie to Mortal Kombat: gender and computer games*. MIT press.

[8] Chickens helping the elderly tackle loneliness - Telegraph: 2014. *http://www.telegraph.co.uk/news/health/11198410/Chickens-helping-the-elderly-tackle-loneliness.html*. Accessed: 2015-06-26.

[9] Connell, R.W. and Messerschmidt, J.W. 2005. Hegemonic masculinity rethinking the concept. *Gender & society*. 19, 6 (2005), 829–859.

[10] Csapo, E. 2014. The Iconography of Comedy. *The Cambridge Companion to Greek Comedy*. M. Revermann, ed. Cambridge University Press. 95–127.

[11] Dill, K.E. and Thill, K.P. 2007. Video game characters and the socialization of gender roles: Young people's perceptions mirror sexist media depictions. *Sex roles*. 57, 11-12 (2007), 851–864.

[12] El-Ibiary, H.M. 1946. The old Egyptian method of incubation. *World's Poultry Science Journal*. 2, 3 (Jan. 1946), 92–98.

[13] Fothergill, B.T. and Sterry 2016. Poulíography and "Poultrymen" in North Africa. *Proceedings of XIe Colloque international Histoire et Archéologie de l'Afrique du Nord* (Marseille et Aix-en-Provence, forthcoming 2016).

[14] Gál, E., Csippán, P., Daróczi-Szabó, L. and Daróczi-Szabó, M. 2010. Evidence of the crested form of domestic hen (Gallus gallus f. domestica) from three post-medieval sites in Hungary. *Journal of Archaeological Science*. 37, 5 (2010), 1065–1072.

[15] Gifford-Gonzalez, D. and Hanotte, O. 2011. Domesticating animals in Africa: Implications of genetic and archaeological findings. *Journal of World Prehistory*. 24, 1 (2011), 1–23.

[16] GTA 5: Grand Theft Chicken - YouTube: 2015. *https://www.youtube.com/watch?v=umEdkaJfLhY*. Accessed: 2015-06-26.

[17] Higham, C. 1989. *The archaeology of mainland Southeast Asia: from 10,000 B.C. to the fall of Angkor*. Cambridge University Press.

[18] Insoll, T. 2010. Talensi animal sacrifice and its archaeological implications. *World Archaeology*. 42, 2 (Jun. 2010), 231–244.

[19] KFC supplier probes poultry abuse | BBC News: 2004. *http://news.bbc.co.uk/1/hi/world/americas/3915599.stm*. Accessed: 2015-06-26.

[20] Kirkland, E. 2009. Masculinity in video games: The gendered gameplay of silent hill. *Camera Obscura*. 24, 2 71 (2009), 161–183.

[21] Lea, S.R.G. 2007. *Delinquency and animal cruelty: myths and realities about social pathology*. LFB Scholarly Pub.

[22] Lentacker, A., Ervynck, A. and Van Neer, W. 2004. The symbolic meaning of the cock. The animal remains from the mithraeum at Tienen. *Roman Mithraism: the Evidence of the Small Finds. Brussels: Instituut voor het Archeologisch Patrimonium*. (2004), 57–80.

[23] Lind, L. 1963. Aldrovandi on chickens: The ornithology of Ulisse Aldrovandi (1600), vol. 2, book 14. (1963).

[24] Livy (Titus Livius). History of Rome. 22.1.18-20. Translated by Aubrey de Selincourt. Edited by Betty Radice (1965).

[25] Macho Man Chickens at Skyrim Nexus - mods and community: 2011. *http://www.nexusmods.com/skyrim/mods/4376/?*. Accessed: 2015-06-26.

[26] Mariano, N. Letter from Nicky Mariano to Derek Hill. Public Record Office of Northern Ireland. Item Reference Number D4400/C/9/10.

[27] Markey, P.M., Markey, C.N. and French, J.E. 2014. Violent Video Games and Real-World Violence: Rhetoric Versus Data. *Psychology of Popular Media Culture*. (2014).

[28] McGovern, T.H., Buckland, P.C., Savory, D., Sveinbjarnardottir, G., Andreason, C. and Skidmore, P. 1983. A Study of the Faunal and Floral Remains from Two Norse Farms in the Western Settlement, Greenland. *Arctic Anthropology*. 20, 2 (1983), 93–120.

[29] PETA says whaling in Assassin's Creed 4 glorifies hurting and killing, Ubisoft responds | Polygon: 2013. *http://www.polygon.com/2013/3/6/4070836/peta-objects-whaling-in-assassins-creed-4*. Accessed: 2015-06-26.

[30] PETA vs. Pokémon: Does The Video Game Teach Animal Cruelty? | The New Republic: 2012. *http://www.newrepublic.com/article/books-and-arts/108500/peta-vs-pokemon-does-the-video-game-teach-animal-cruelty*. Accessed: 2015-06-26.

[31] plorry comments on Animal cruelty and video games: 2014. *http://www.reddit.com/r/vegan/comments/2l5j3z/animal_cruelty_and_video_games/clrqbh5*. Accessed: 2015-06-26.

[32] Pratchett, T. 1991. *Witches abroad*. V. Gollancz.

[33] Procopius. The Vandalic War. (III.2.25–26)

[34] Ruhlman, M. and Turner, D. 2014. *Egg: a culinary exploration of the world's most versatile ingredient*. Little Brown & Co.

[35] Sayer, K. 2013. His Footmarks on Her Shoulders: the Significance and Place of Women within Poultry Keeping in the British Countryside, c. 1880-c. 1970. *Agricultural History review*. 61, II (2013), 301–29.

[36] Scharrer, E. 2005. Hypermasculinity, aggression, and television violence: An experiment. *Media Psychology*. 7, 4 (2005), 353–376.

[37] Shoemaker, R.B. 2007. *The London mob violence and disorder in eighteenth-century England*. Hambledon Continuum.

[38] Simpson, J. and Roud, S. 2000. *A dictionary of English folklore*. Oxford University Press.

[39] Singer, P. 1975. Down on the factory farm. *Animal Liberation. A New Ethics for Our Treatment of Animals. New York: Avon Books. The Hearst Corporation*. (1975), 159–68.

[40] Stirling, L.M. 2004. Archaeological Evidence for Food Offerings in the Graves of Roman North Africa. *Daimonopylai: Essays in Classics and the Classical Tradition presented to Edmund G. Berry*. R.B. Egan and M.A. Joyal, eds. University of Manitoba Centre for Hellenic Civilization. 427–51.

[41] Storey, A.A. et al. 2012. Investigating the Global Dispersal of Chickens in Prehistory Using Ancient Mitochondrial DNA Signatures. *PLoS ONE*. 7, 7 (Jul. 2012), e39171.

[42] Strutt, J. 1801. *The sports and pastimes of the people of England : including the rural and domestic recreations, May games, mummeries, shows, processions, pageants, and pompous spectacles, from the earliest period to the present time*. William Reeves.

[43] Sul, H. 2000. The King's Book of Sports: The Nature of Leisure in Early Modern England. *The International Journal of the History of Sport*. 17, 4 (2000), 167–179.

[44] Sykes, N. 2012. A social perspective on the introduction of exotic animals: the case of the chicken. *World Archaeology*. 44, 1 (Mar. 2012), 158–169.

[45] Video Games Normalize Animal Cruelty | CounterPunch: 2013. *http://www.counterpunch.org/2013/11/29/video-games-normalize-animal-cruelty/*. Accessed: 2015-06-26.

[46] Wells, C. 2008. The Mystery of Socrates' Last Words. *Arion*. (2008), 137–148.

[47] *The Knickerbocker* 1847. March 1847, 283. Available at http://bit.ly/1Gwgepx. Accessed: 2015-06-26

10. APPENDIX: Video Games Reviewed

Game	Year	Game	Year	Game	Year
Freeway	1981	Fable: Anniversary	2004	Gears of War 3	2011
Chuckie Egg	1983	Grand Theft Auto series	2004	Orcs Must Die	2011
Legend of Zelda	1986+	Sly 3: Honour Among Thieves	2005	Guild Wars 2	2012
Castlevania Series	1986+	Resident Evil 4	2005	Puzzle Craft	2012
Final Fantasy Series	1987+	Fable: The Lost Chapters	2005	Farming Simulator	2012
Monkey Island	1990	Hit Man: Blood Money	2006	Resident Evil 6	2012
Pokemon	1996	Vanguard: Saga of Heroes	2007	Dust: An Elysian Tail	2012
Harvest Moon	1996+	Crysis	2007	Diablo 3	2012
Tekken 3	1997	Portal	2007	Counterstrike: Global Offensive	2012
Gauntlet Legends	1998	Lord of the Rings Online	2007	Far Cry 3	2012
Mort the Chicken	2000	Fable 2	2008	Guacamelee	2013
Banjo-tooie	2000	Minecraft	2009	Legend of Dungeon	2013
Animal Crossing	2001+	Resident Evil 5	2009	Far Cry 4	2014
EverQuest 2	2002	Monsiegneur Cockburn (Doom clone)	2009	Forge Quest	2014
Billy Hatcher and the Giant Egg	2003	Fable 3	2010	DayZ Standalone	2014
World of Warcraft	2004	Rift	2011	Besiege	2015
Fable	2004	Skyrim	2011	Metal Gear Solid V	2015
		Witcher 2	2011	Divinity: Original Sin	2015

Digital Alienation as the Foundation of Online Privacy Concerns

Brandt Dainow
Department of Computer Science, Maynooth University
Kildare, Co. Kildare
Ireland
+353 86 248 2846
brandt.dainow@nuim.ie

ABSTRACT

The term 'digital alienation' is used in critical IS research to refer to manifestations of alienation online. This paper explores the difficulties of using a traditional Marxist analysis to account for digital alienation. The problem is that the activity people undertake online does not look coerced or estranged from the creator's individuality, both of which are typically seen as necessary for the production of alienation. As a result of this apparent difficulty, much of the research has focused on the relationship between digital alienation and digital labour.

This paper attempts to overcome these difficulties by discarding the traditional approach. We argue one can better understand digital alienation by focusing on the relationship between user intent and technical infrastructure, rather than concerns with labour. Under the existing economic model dominating the internet, free services are financed by recording user activity and then using the products of this commercial surveillance to sell information about people to others. We show how the real harm in current online business models is that commercial surveillance is being used to commodify private life.

Seeking to define personal data in more precise terms, we will introduce two new concepts necessary for a detailed discussion of any ethical issues regarding personal data - the digital shadow and the digital persona. We will then show how affordances in current online systems are tuned to commodification of the user's personality. We will then explore the nature of online surveillance and show how affordances combine with the surveillance economy to produce digital alienation.

Categories and Subject Descriptors

K.4.1 **[Computers and Society]:** Public Policy Issues: Ethics, Privacy

General Terms

Management, Economics, Human Factors.

Keywords

digital alienation, privacy, digital economy, surveillance, targeted advertising, personalization, critical theory, ICT ethics, Marxism

Conference'10, Month 1–2, 2010, City, State, Country.

1. INTRODUCTION

Digital alienation is a privacy issue. Digital alienation occurs when one's digital lifeworld or the digital self is exploited. The process of exploitation extracts value from a person's digital activity through coercion and manipulation. We are coerced into submission to ubiquitous commercial surveillance[1] of our digital activity. Value is extracted from this surveillance process through the conversion of surveilled data into economic and political capital. The entire system represents a reification of one's digital lifeworld and commodification of the digital self. It also poses a number of problems for traditional understandings of concepts related to alienation within Marxist theory, such as coercion, exploitation, and power dynamics. Indeed, much of the debate in this area over the last few years has been concerned with how to account for alienation within a digital context. I believe the solutions to current problems can best be achieved by altering the analytic approach.

2. SCOPE OF CONCERN

Being connected to the ubiquitous computing environment which is coming to surround us is already necessary for full participation in modern Western societies. A review across the range of those emerging ICT's which will impact society over the next decade shows that being connected may become necessary for survival itself [35]. When the internet first emerged, it was predicted that it would "flatten" the power structures of traditional society, even lead to the "fading away" of the nation state [57]. Such views were based on technological determinism; they envisioned the new distinguishing features of internet technology as passing unmodified into society and reshaping it to match the internet's technical architecture [15].

In reality, the development of the internet ecosystem has been filtered through the structures of pre-existing society and evolved in accordance with its imperatives. While it has been disruptive in terms of changing some of the dominant players in media markets, destroying some and creating others, it has not fundamentally changed the power structures in society. Authoritarian governments have learned to control and censor it, hegemonic corporate capitalism has come to dominate it, and people's digital activities have been cajoled into closed silos controlled by a very few exceptionally large corporations [15]. Once seen as the antidote to structural inequality, the internet has actually become a profoundly powerful tool of domination based on exploitation and alienation.

3. ALIENATION IN CRITICAL IS STUDIES

The term 'digital alienation' is used in Critical IS research to refer to manifestations of alienation online. Stemming from

[1] Commercial surveillance involves the recording and analysis of online user behaviour with the aim of predicting and controlling their behaviour [81].

digital labour studies [43] the focus soon bridged into social networking. A good example of this bridging can be seen in Fuchs and Sandoval's *Framework for Critically Theorising and Analysing Digital Labour* [25]. Initially exploring the dimensions of paid digital labour, the authors extend the analysis into the realm of unpaid labour within content production in social networks. P.J. Rey's paper *Alienation, Exploitation and Social Media* [66] explores the mechanisms by which capitalism has come to exploit social media. Rey's task involves demonstrating how alienation exists within social networking as a dynamic of value extraction. This approach is also used by Christian Fuchs [23,25] in most of his work. By contrast, Krüger and Johanssen's *Alienation and Digital Labour—A Depth-Hermeneutic Inquiry* [43] examines alienation through a survey of prosumer's comments about social network's themselves. Here alienation is demonstrated through the effects of the social network system's activities, rather than through the dynamics of labour and surplus value extraction. If alienation can derive from unpaid digital labour, as seen in social networks, the possibility arises that alienation can be found wherever unpaid digital labour occurs. Here we find Marc Andrejevic's *Surveillance and Alienation in the Online Economy* [6], which extends the analysis of alienation beyond social networking into general online activity. This paper shows a third approach to explaining digital alienation by focusing on exploitation, in contrast to the previously mentioned papers, which focus on value extraction and coercion. What all these analyses demonstrate is that the nature of alienation online necessarily diverges from the account of alienation in earlier, pre-digital, analyses. These divergences reflect the differences in structures of production and value-extraction between analogue and digital socio-technical systems. These differences are significant to the degree we may warrantably talk of a distinct "digital" form of alienation.

In Marx, alienation is the result of labour activity coerced into alienated forms in order to produce products estranged from the producer [60]. The political dynamic is the extraction of value from controlled and structured worker activity. Historically, analysis of digital alienation has focused on accounting for the traditional mechanisms underpinning alienation within a digital context. There has been an unspoken consensus that an account of digital alienation requires identifying the same structures and mechanisms within the digital context as Marx identified within the factory. Here the concern is to understand digital alienation by analysing it as the result of conditions considered necessary for alienation - coercion, labour and estrangement from product. With regard to coercion, the difficulty is whether people who freely choose to use social networks like Facebook can be described as coerced. The concern with labour is whether people's unpaid production of content in social networks can be described as labour. Finally, if people seem to be expressing themselves within social networks, the question arises as to how can they estranged from the output of their activity.

At one extreme researchers such as P.J. Rey have argued that the differences between the Victorian factory of Marx's analysis and modern digital activity are so great that alienation is a questionable concept within a digital context [66]. Rey argues that the products of digital labour in social networks are not alienated because creation of this content is freely chosen and creative. Referring back to Marx's categorisation of imagination as a distinguishing characteristic between animals and humans, Rey suggests that the creative nature of social content production renders the output unalienated. His view is that the creativity involved in social network content creation allows the producer to recognise themselves within their output. In addition, the free choice to engage in social networking means this labour is uncoerced. Rey does accept there is some degree of exploitation involved because social networks derive financial value from this output without financially compensating those who produced it. However, he argues this exploitation is mild because producers do receive compensation in other forms of capital. Rey argues that social network users are compensated because they retain use of their output for their own purposes. They can therefore use the content they produce to generate social and cultural capital. His position is that the non-economic value derived is so great that any exploitation is "relatively minimal" [66:415]. Furthermore, any exploitation present is, Rey argues, further diminished by the unalienated nature of prosumer output. Rey acknowledges that social networks also derive value from surveillance of user activity, usually without users being aware of it. However, while he sees this as mildly exploitative, he does not consider it alienating. Rey's position is that digital capitalism can maintain the inequalities and power structures within society identified by Marx, but without the need for alienation, or even very much exploitation.

It is notable that, while recognising that social networks extract value from user surveillance, Rey does not extend this recognition to the fact, noted by others, that such surveillance is almost universal throughout the internet [51,81]. A 2012 study of the world's busiest websites revealed that 94% engaged in some form of user surveillance themselves, half of whom also allowed unidentified third parties to engage in such tracking through their sites. The same study also found that 91% of these sites changed their content to match their understanding of the user [70], something impossible without a pre-existing knowledge of that user; knowledge which can only have come from previous surveillance. User activity in other parts of the internet, such as search and reading, does not generate cultural or social capital, but is still subject to the same levels of commercial surveillance. Following the logic of Rey's analysis, this renders such surveillance much more exploitative. In general, Rey's analysis treats technology as invisible and as permitting users to fully express themselves in an unmediated fashion. While Rey recognises that surveillance occurs, he fails to take into account that much it is used to tune and filter the online environment surrounding the user. Users are presented with "personalised" choices, links and content based on the results of covert surveillance as much as on the content they produce; something often referred to as the "filter bubble" [56,65]. People are therefore not able to make free choices or even fully express themselves, because the technology available to them is not value-neutral, but tuned to commodification [19,39].

Andrejevic's analysis of digital alienation is founded on just this consideration. All internet users are subject to pervasive universal surveillance by commercial enterprises [16,70,79,81]. The value of this surveillance far exceeds that derived from social network content creation [16,20,29,51,79,81]. Initially this information was used only to tune advertising delivery [16,81]. However, this information is now also used to tune the delivery of news on many sites [81] and for political manipulation [1]. Users have no choice over whether their activity online is recorded, processed and used, nor do they know who by [70]. This constitutes, for Andrejevic, alienation. His argument is that the lack of choice over whether to be surveilled or not constitutes a structurally-embedded coercion. He further argues that the lack of knowledge about this surveillance constitutes an epistemological alienation. Finally, he argues that the use of this information to alter content in an effort to manipulate the user fits Marx's definition of alienation as an estranged power structure working against the individual [6,8].

In contrast to Rey's position that exploitation is mild because the user derives non-economic use-value from the content they create, Christian Fuchs [23] has argued that exploitation is either present or not, and cannot be present in variable degrees. One cannot be a little bit exploited. Fuch's work tends to focus on the mechanisms of value-extraction within a digital context. Fuch's

position is that any activity conducted by someone which can be used to generate economic value is labour. He seeks to bring together the competing positions held by Andrejevic and Rey by arguing that Rey is focused on subjective feelings of alienation whereas Andrejevic is focused on the objective conditions of non-control and non-ownership. However, Fuchs firmly comes down on the side of alienation being objectively present, arguing that the purported social use-value that content creators derive from their work hides the true commodity character of social networking [24]. He identifies two dimensions of value within social networks - the value of created content and the value of user presence. Here Fuchs agrees with Andrejevic that the users of social networks are themselves treated as commodified products which are then sold.

Both Fuchs and Andrejevic limit their conception of the use of personal data to the realm of advertising delivery. While the first use of this information was indeed to tune content, especially advertisements, to the user profile, this information is now also sold for other purposes, including political manipulation [1], credit scoring [54,72], housing and employment [12] and news delivery [81]. It is worth noting that both Facebook and the international trade body for online advertising, the Internet Advertising Bureau, agree with this assessment of where the real value lies in commercial online surveillance [16,20]. In comparison with this vast and pervasive surveillance industry, user-generated content within social networks is a trivial consideration. Under this analysis, alienation is a pervasive and unavoidable adjunct to almost all digital activity.

Rey, Andrejevic and Fuchs all approach alienation within a digital environment by focusing on Marx's mechanisms for its production and explaining how and where these mechanisms can be found online. While general commercial surveillance is mentioned, it is not really the central focus of their analysis, nor does it alter their approach. My position is that we can better account for digital alienation if we can liberate ourselves from the form of Marx's account. Marx provided an analysis of how alienation occurred within a particular historical and technological context. As we have seen from the above, we encounter problems if we assume that this is the only mechanism by which alienation can occur or that all of these traditional mechanisms are necessary. My argument is that the features of digital alienation are so different from traditional alienation that a new account is necessary.

4. HOW ALIENATION OCCURS ONLINE

In defining alienation, Marx considered two factors, the nature of alienation and the means by which it is produced. The nature of alienation is that the individual is disconnected from the products of their labour by property ownership rights; they are alienated from ownership of both the product and the means of producing it. This constitutes the material base of alienation and is the product of power relations governing the production process. Marx's account involved material coercion by controlling access to the means of survival so as to force people into alienating labour. Analysis of exploitation on the internet has been distracted by the apparent lack of coercion motivating online activity and by the appearance of self-expression in social networks. However, our analysis becomes less complicated if treat social networking within the broader context of pervasive digital surveillance. Here we recognise that, while content production in social networks is voluntary and can be self-expressive, it is just one type of action within the wider class of voluntary and self-serving digital activity which includes search, shopping, email, use of maps, health trackers, life loggers and other digital services, not to mention general web surfing. This is important because the range of digital activities will continue to spread until it permeates most of our environment [63]. Because of this it is essential to treat the current state of affairs as an intermediate process moving towards more ubiquitous computing. Our analysis must recognise that the political and economic structures which affect us within the current digital domain are on a trajectory to dominate our entire existence, offline as well as online. It is important, therefore, to recognise that the frame of analysis cannot limit itself to voluntary activity knowingly making use of digital services. The infrastructure being created now will one day support smart cities, the internet of things, and digital devices implanted within our bodies. Our entire existence will become mediated through digital services within a few decades [63].

Thus the place of labour as seen in a traditional account of alienation becomes problematic when value is being extracted from broad-spectrum use of digital services for life in general. Assuming that labour is a necessary precondition for alienation requires explaining how all activity using digital services constitutes labour despite the fact it generates no obvious income and may not even be anything more than a traditional activity, like walking or driving, which has been supplemented with a digital component. Certainly the argument of remuneration in the form of social or cultural capital is inapplicable with reference to activities which do not involve any form of communication, such as using search engines or passively reading a website, yet value is extracted from these activities by others via commercial surveillance [8,11,16,62,69,73,81]. If we redefine 'labour' as referring to any activity from which value may be drawn by any party, as Fuchs does [23,24], then almost all activity becomes labour and the term ceases to provide any real distinction from other mode of activity. I think it is better to abandon the issue of whether online activity is labour or not. There is nothing within Marx's description of alienation which requires that it must, of necessity, derive from labour. 'Alienation' in Marx is not a single concept, but a translation of two terms, *Entfremdung* and *Entäusserung*, which can also be translated as 'estrangement' and 'externalization' respectively [60]. These terms are applied to a variety of phenomena, including internal mental states, property relations and societal structures. It is true that Marx attempts to provide a systematic analysis of political economy based on the concept of alienated labour in his early work, but that attempt is incomplete [86]. In his later works, alienation becomes a descriptive term which is applied to multiple phenomena. There is nothing in his usage which locks alienation to labour except as a historically contingent feature of nineteenth century capitalism [86]. All that is required by Marx's account is that there be human activity and that this occur within certain types of unequal power structure within the field of economic competition.

On this basis, I propose to focus on digital alienation as a product of property relations regarding data. Surveillance is a process of data acquisition; some generated as the output of online surveillance monitoring systems and some data taken from elsewhere, such as the passenger name records used for international travel, geo-location data and credit scores [1]. The common element all these data elements have is that they are held to pertain to the same individual[2]. The dataset created is termed a "personal profile", as opposed to group profiles [31,38]. The personal profile is a digital representation of an individual. It is the central commodity of the surveillance economy. Each organisation which holds a personal profile subjects it to algorithmic analysis and manipulation in order to extract value from it. The term used in the industry is to "monetize" it. It is this profile which is used to tune content and for purposes of manipulation. All actions using personal data draw that data from the personal profile. Such use constitutes Marx's concept of an environment which reflects back on the producer estranged output

[2] This belief may be mistaken, it is not always possible to distinguish between a person and a device; and cases of mistaken identity also occur.

[8]. In that surveillance technology produces the personal profile as a commodity, it is a type of production process. The raw material for this production process is the activity of individuals [32], which is used to produce personal profiles. This production process is not owned by those who generate the activity which feeds it. This is the basis for alienation from ownership of the means of production. The surveillance process is hidden and unwelcome [14,32] and therefore represents an unequal, coercive and exploitative power structure [8]. We may view the personal profile as a field of contention between commercial surveillance companies and those who use their products on one hand opposed by individuals and privacy advocates on the other.

The essential starting point to all forms of alienation is individual activity. I therefore believe we may best understand digital alienation by examining the mechanisms by which an individual's digital activity is alienated. Here we must focus on the nature of personal action within a digital context, the mechanisms by which the personal profile is generated, and the use to which it is put. As mentioned above, the first task is to dispense with the need for a concept of labour. In Marx's analysis labour was the term used to distinguish activity which supported alienation from activity which did not. Labour supported alienation because it was activity which occurred within, and was shaped by, exploitative power structures. However, no such distinction between labour and non-labour exists online because all activity is surveilled and exploited [73,81]. Not only does discarding the need for labour ease our analysis, I believe it helps to direct our attention to the ubiquity of digital surveillance. Instead I will define human activity within a digital context in terms of people's intentions and expectations. To do this I will introduce distinguish the two targets of surveillance; communicative activity and everything else. I will refer to these as the 'digital persona' and the 'data shadow', respectively.

'Digital persona' is the term I propose for the body of digital material created by an individual through acts of online communication. The digital persona includes blogs, comments, product reviews, tweets and other social network postings, together with any other conscious communication by an individual within a digital context. Thus the digital persona is created by the individual to express and communicate. The digital persona is not a direct or unmediated reflection of the personality, but a creation through which the individual seeks to represent of an aspect of themselves. The disconnect between the offline and digital world permits people to exaggerate or repress particular aspects of their personality [77]. For example, introverts may use the digital persona to compensate for difficulties they have in face-to-face interactions [3] while extroverts often use it to confirm pre-existing characteristics [82]. In other cases, people develop new personal characteristics online so that they can incorporate them into their offline personality [53]. In all cases, what is revealed or portrayed is further influenced by previous experiences online, especially concerns over privacy and security [42,87]. I derive the term 'persona' from C. G. Jung's concept of the persona as a creation of the ego designed to represent a subset of that ego within specific social circumstances [36]. The same idea is used within a sociological perspective in Goffman's *The Presentation of Self in Everyday Life* under the term 'masks' [28], which outlines his Dramaturgical Theory, a sub-set of Symbolic Interactionism [68].

The term 'data shadow' was first used by Alan Westin in *Privacy and Freedom* [84], but has entered into general use both in computing and privacy discussions. It refers to the information generated by someone as a side-effect of their use of digital technology. These days this includes log files, access records, search histories, movements between and within web sites, mobile phone location records and all financial activities not involving cash [11,31,39]. Thus the term 'data shadow' refers to all digital information pertaining to an individual which they did not consciously and intentionally create for communicative purposes. This information may have been generated by the user for other purposes, such as their "click-stream history," which is a record of their mouse click activity within a website [32], or their "search history," a record of all the searches they have made in a given search engine. Elements of the data shadow can also be generated through the monitoring and recording of user activity by other systems. For example, web server log files, containing records of every file request, constitute data generated by the system about the user. The term 'data shadow' includes the material used to commodify users within social networks, but also applies outside social networking. Data shadows may be created through any and all use of digital technology.

Data shadows are created by a network of commercial surveillance agencies whose tracking technologies permeate digital services [11,16,46,51,81]. Very few of these agencies are known to the public [11,73]. Some, like Google and Facebook, are well known because of their public profile as digital service providers, though their activity as commercial surveillance agents is less well known, even though it drives their profits [20,29]. Others, such as DoubleClick, Acxiom, Experian and BlueKai are known to industry analysts and privacy advocates as a result of their scale and reach. However, the majority, such as ClickTale, Optimzly, Kiss Metrics, Info Group, Ace Metrics, Crazy Egg, Site Meter, Moz, Adgistics, People Metrics, Data Dog, Data Mentors, Extrawatch, Inspectlet, eDataSource, Prognoz, and literally hundreds of others, are unknown outside the specialist profiling industry. No one knows how many of these agencies there are, or what they do, but it is known they combine the data they gather with information from other sources to create detailed profiles on literally hundreds of millions, if not billions, of people [11,21,83]. The commercial surveillance industry is much larger in terms of economic value and user-base than any other online industry [16,73,81].

This universal commercial surveillance means there is no way to use most digital services without being surveilled [21,32,73,81,83]. For most digital services there is no alternative provider who does not practice surveillance (or permit others to do so) within the service stream [76,81]. However, lack of choice most strongly stems from lack of knowledge. We are simply unaware of when we are being surveilled, who by and for what purpose [32,79]. Obviously, one cannot exercise choice over things one is unaware of. As we have seen, this lack of choice has been held to constitute coercion by Andrejevic and Fuchs, but not by Rey. Lack of choice as coercion has a long history of support in philosophy. For example, Aquinas argues that coercion occurs when actions by one person mean someone cannot act otherwise [10]. However, this position was challenged in the twentieth century by the position that coercion requires communication between the coercer and their target, usually in the form of conditional threats [2]. Under this view coercion is a communicative act, not a contextualising situation. This is the position currently supported in much legal practice, especially in the USA [5]. However, since the 1980's arguments have re-emerged in support of structural coercion; the creation of situations in which one is prevented from selecting alternative courses of action [67]. Here the focus is shifted to the coercer's intentions to remove choice from another [4]. This accords with much of Marx's analysis in which he focuses on the general circumstances of capitalist society as coercive in the sense of removing freedom [86]. Clearly, hiding surveillance so that people cannot avoid it constitutes removal of choice and diminution of freedom. Thus it is possible to argue from this perspective that the lack of choice to avoid surveillance constitutes coercion. However, we must recognise that this position is not in accord with how many, especially in jurisprudence, understand the term.

Lack of choice, even coercion, does not automatically mean that the output of a productive process is alienated. I wish therefore to explore the mechanism by which digital activity becomes alienated. Since we have two forms of digital data, the digital persona and the data shadow, two accounts are necessary. I shall commence with the alienation of the digital persona.

5. EGO, AFFORDANCES AND THE DIGITAL PERSONA

People use Web 2.0 technologies to create their digital persona. The process by which they do this, and the persona they create, are alienated. We therefore need an account of the mechanism by which people do this and how alienation occurs. Central to my account of how the digital persona is alienated is the view of technology as a socio-technical system [35]. A technology may be composed of multiple artefacts and may be "read" or understood in different ways [30,34]. The nature of the "reading" depends on the person, their social environment, past experiences and other factors, all of which are constrained by the functional capabilities of the artefacts in question [58]. We therefore need a conceptual framework which holds all the dynamics which are at play in a person's understanding and use of a technical system. I will use the concept of "affordances" to explain the interaction between people and the technical artefacts.

The concept of affordances originates with James Gibson's conceptualisation on the subject of how animals perceive and understand their environment in *The Ecological Approach to Visual Perception* [27]

> "The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill… It implies the complementarity of the animal and the environment… [affordances] have to be measured relative to the animal. They are unique for that animal. They are not just abstract physical properties." [27:127]

Gibson was arguing against a reductionist understanding of perception and for a perceptive process within all animals in which perception itself is not merely a process of physical activity onto which understanding is overlaid *post hoc*. Instead, he argued that the perceptive process itself incorporates cognitive elements such as motivation, environmental context and past experience into the act of seeing.

The concept was applied to ICT analysis by Ian Hutchby in *Technologies, Texts and Affordances* [34], in which he describes technologies as

> "Texts which are written in certain ways by their developers, producers and marketers, and have to be read by their users or consumers. The writers of these technology texts may seek to impose particular meanings on the artefact, and to constrain the range of possible interpretations open to users. Users, by contrast, may seek to produce readings of the technology texts which best suit the purposes they have in mind for the artefact… Neither the writing or reading of technology texts is determinate: both are open, negotiated processes. Although there may be ways that technology texts have preferred readings built into them, it is always open to the user to find a way around this attempt at interpretive closure." [34:445]

We may thus see affordances as a field of competition in which the owners of a technology compete with the users of that technology for domination of the affordances dictating how that technology is understood and used. Donald Norman explores this competition over technological affordances in *The Design of Everyday Things* [58]. In Norman's account, we use affordances to build conceptual models of how things work. Any technology involves the interaction of two conceptual models; a design model and a user model. The design model is the conceptual model held by the designers when they built the technology and in accord with which they try to construct the artefact. The user's model is the conceptual model users have of that same technology. Norman is concerned with what happens when the two models clash or diverge. According to Norman, there is no necessary convergence between the user's mental model and the designer's. In fact, in Norman's view, the two model's clash most of the time. Using Norman's framework, I suggest that the user model conceptualises the Web 2.0 services people use to express their digital personas as private, unmediated and natural. The user model fails to recognise the degree of surveillance and the degree to which their activities are mediated through a technology designed for data gathering and commodification. Users also fail to recognise the degree to which surveillance is used to filter and control the content they see in social networking and news sites and in advertising. Instead, users see the content presented to them within social networks as somehow neutral, unmediated and unsurveilled [71]. In contrast, service providers, such as Google and Facebook, show evidence of believing that users have the same conceptual model as designers. They have countered concerns over online privacy by stating users have no expectation of privacy and accept that the material they create will be processed for purposes of commodification [22].

There are numerous studies which demonstrate that users manipulate their self-expression online in order to convey specific characteristics and control the image others have of them [41,50,53,82]. In our terminology we may say people use Web 2.0 technologies to construct their digital personas. Their understanding of what can be expressed, the values determining what should be expressed and how this is to be done are determined by the affordances users perceive in these technologies [74,89]. These affordances constitute what Groffman describes as the "props and tasks" [28:143] which dictate what persona[3] is appropriate and the "expressive resources" [55:74] available from which to construct it.

Unfortunately for users, Facebook and similar Web 2.0 systems are not designed for people to portray themselves in any manner they may choose. Instead, Facebook and similar systems divide personal characteristics into a set of discrete data points, such as preferred objects of consumption, marketable skills, and approvable attitudes [33]. Furthermore, qualitative characteristics, such as friendship, are reduced to quantitative values, such as the number of likes or followers. Facebook's affordances, in particular, suggest to users that their digital persona is a true reflection of their identity, yet is at the same time something to be constructed, managed and enhanced [26]. Facebook openly expresses the neo-liberal concept of a "personal brand," in which a person creates a commodified public image as the repository of their social capital [44]. The affordances of Facebook present the individual as composed of consumption patterns (such as preferred movies, books and music) and patterns of association (as shown through one's likes, friends and photos). These are dimensions of analysis more suited to processing for advertising than developing an understanding of the whole person. There is good empirical evidence that this model conflicts with the affordances the user brings to Facebook. In many cases users seek to express themselves in ways restricted by the affordances Facebook imposes, resulting in dissatisfaction, resistance and disuse [26,42,50,74].

In using affordances tuned to atomising, quantifying and commodifying the depiction of people, social media systems like Facebook alienate the digital persona. Rather than a free expression of the self, users are forced to display only those

[3] Groffman's term is "performance" [28:143]

characteristics which are commodifiable. These characteristics are then embedded in a manipulated content environment which reinforces and promotes ongoing commodification, and therefore embeds the alienated digital persona within an alienated social environment.

6. ALIENATION AND THE DATA SHADOW

The mechanisms by which the data shadow is alienated are straightforward compared to the digital persona and commence with unavoidable, hidden, ubiquitous commercial surveillance [70,73,81]. In that the digital world is permeated with unknown entities gathering unknown information to use for unknown purposes [21,83], the digital environment is self-evidently epistemologically alienated from the user. The material base of the commercial surveillance system supports a superstructure devoted to exerting power over the individual by influencing their behaviour directly, or by influencing decisions made about them by other people [11,72,80]. This is achieved through personalization of content [56], such as the advertising [78] and news [81] to which people are exposed.

The unavoidability of commercial surveillance is made possible by the lack of ownership or control users have over digital services. It is known that, in general, people do not like commercial surveillance or content personalization [14,17]. Commercial surveillance therefore constitutes the exercise of power over individuals and a diminution of their freedom, another manifestation of alienation [7]. Furthermore, the knowledge that unknown surveillance is occurring, in combination with lack of knowledge about how that information is used, has a chilling effect on people's online activities [32,49,75]. In effect, people are alienated from their own actions online before they perform them. In that this chilling effect also applies to how people communicate online, ubiquitous commercial surveillance further alienates people from each other.

7. DIGITAL ALIENATION – THE COMPLETE PICTURE

We are now in a position to provide an account of how the four dimensions of alienation occur. First, users are alienated from their productive activity through restricted affordances within expressive Web 2.0 technologies which promote a commodity fetishism of personal characteristics and interpersonal relationships. This is made possible by an alienated power structure which is designed around treating users as commodities [8,23]. Users are alienated from non-expressive activity by the presence of ubiquitous hidden surveillance systems. Thus users are alienated from all forms of digital activity. Second, users are alienated from the products of their digital activity by property relations. These grant service owners the right to reuse user-produced content for their own purposes and to process both the digital persona and the data shadow in order to construct personal profiles. Users are further alienated from the products of their own activity since the personal profile is used against them, either to manipulate their behaviour or to influence how others treat them. In addition, the abandonment of the open standards which created the web means that the products of user activity are imprisoned within data silos owned by service providers [18]. Thus, you may close your Facebook account, but you can't move it to another social network. Third, users are alienated from each other by the necessary mediation of fetishizing social networks and by the chilling effect of ubiquitous surveillance. Finally, users are alienated from themselves and their own human potential in three ways; through the imposition of fetishizing affordances promoting the concept of the personal brand, through their limited control over their own digital persona, and through the use of personalization technologies which confine the user's ability to discover the unexpected, the unusual, and the uncommodified.

8. SOLUTIONS

No solution exists today which can resist these patterns and structures of digital alienation. However, a number of technologies exist which can form part of a solution, while the design principles to complete the solution are understood. Two related characteristics support the existence of digital alienation, lack of choice and lack of power. The solution is therefore to restore choice and empower the user. In my view solutions that look to regulation, such as data protection and privacy laws, merely perpetuate a hierarchical structure which keeps people in a powerless position. Instead of companies deciding what to surveil, we merely pass the decision to legislators. Given the history of government digital surveillance [9] there is nothing to suggest this improves matters. In addition, the impossibility of a single legislative framework for the entire internet [64,85,88] means surveillance companies can simply move to more conducive regimes. Furthermore, centralised storage of personal data is frequently subject to leaks [40,47,48], so I am opposed to centralised storage of any fashion, never mind under what rules.

The first task in combating alienation must be to remove coercion from the situation by giving users the choice over whether to be surveilled and for what purpose. A number of technologies exist which can offer elements of this solution. Anonymizing systems such as such as TOR [52] and TextSecure [61] enable users to avoid being tracked while using the existing internet. These need to be extended and built into a comprehensive set of easy-to-use systems which can wrap browsers and other applications in a protective and intelligent layer which negotiates and controls what data is accessed by what services. Protocols like the W3C's Platform for Privacy Preferences (PPP) [13] can form the basis for such communications. Design of data gathering systems should follow principles of privacy preservation, such as those developed by Marc Langheinrich [45], one of the authors of PPP. Such technology would enable users to control how much information is gathered about them and thus how much personalization is possible. This de-alienates the productive technology by putting control in the hands of the user and de-alienates their digital environment by permitting them to control or prevent personalization.

While these solutions restore choice to the user, they only partially redress the balance in an existing system which is structurally inequitable. The long-term solution must therefore be to move personal data storage, and therefore ownership, into the hands of the users. Here the solution is to reverse the cloud architecture. Currently, centralised systems run analyses of locally held data. I propose inverting this structure, such that personal data is held by the person in their own devices. Effectively each person, or home, would operate their own data store. Following Langheinrich's privacy-preserving design principles [45], devices would, wherever feasible, store their own data. A personal server or gateway would provide the interface between digital service providers and the user's personal data. This gateway would be able to negotiate access for services and prepare personal data for access. This pre-processing would anonymise the data to the degree selected by the user for that type of service. I envision this system working in a manner similar to hierarchical protection domains (or "security rings") within chipsets. These create a series of layers within which particular software operations can be confined so as to shield the system from inappropriate operations [37]. Corporate digital services could still be centrally managed and owned, but their computations would have to call on the individual's own data store rather than house it on corporate servers.

This is, however, merely a collection of artefacts. As a socially-embedded system, technology needs more than just hardware if it is to be adopted. The additional component required is therefore societal structures promoting and maintaining such a system. A network of local technicians is required to maintain and develop such systems, provide advice and training, lobby regulators for support and so forth. Here I suggest the basis lies in recognising the value of personal data. For example, the value of a Facebook user is between $US40 and $US300 [59]. If personal data has value for service providers, let them pay for it. A system of micropayments for access to personal data would create a data economy enabling individuals to earn money through the gathering and storing of their own data. Support agencies, such as technical staff and software vendors, can then be remunerated through a share of this income. Such a system would permit the development of an intermediate layer of data vendors who can store and provide personal data on the user's behalf, according to guidelines provided by those users, or remotely maintain data held in the home. Such a system permits of multiple organisational models. Community groups could operate such services. For example, people who share the same set of data access protocols could form cooperatives to manage storage and access to their member's data. As yet, such technology does not exist. However, the hardware is already in place. Personal cloud storage devices have been available for several years. These permit users to store their data in their home while still being able to access it remotely. The missing components are therefore the micropayment and data negotiation systems. Protocols exist which can handle both, they merely need to be implemented as working products.

We need to bear in mind that the digital service infrastructure we see today is merely a step towards a digital environment of ubiquitous devices; embedded within our bodies, throughout our homes, offices, cars and public spaces. A critical evaluation of current data practices must consider this long-term future and seek emancipatory paths within it. As we have seen, digital alienation is the product primarily of inequitable power structures which intentionally deny users control, or even knowledge, of what is being done to them. The motive power of these structures is the economic value of personal data. If digital services are to align with individual needs, we cannot avoid personal data being processed. The solution is therefore to develop systems which pass some of that value back to the user. Doing so gives the user power and makes them a viable partner for other organisations who can earn a living by controlling access to personal data on behalf of the user. Giving the individual control over their personal data emancipates them from subjection to hegemonic digital capitalism by permitting them to negotiate the terms of the relationship they have with their digital service providers.

8. REFERENCES

[1] Nathan Abse. 2012. *Microtargeted Political Advertising in Election 2012*. Internet Advertising Bureau, New York, NY, USA. Retrieved from http://www.iab.net/media/file/Innovations_In_Web_Marketing_and_Advertising_delivery.pdf

[2] Timo Airaksinen. 1988. An Analysis of Coercion. *Journal of Peace Research* 25, 3, pp. 213–227.

[3] Yair Amichai-Hamburger, Galit Wainapel, and Shaul Fox. 2002. "On the Internet No One Knows I'm an Introvert": Extroversion, Neuroticism, and Internet Interaction. *CyberPsychology & Behavior* 5, 2, 125–128. http://doi.org/10.1089/109493102753770507

[4] Scott Anderson. 2008. Of Theories of Coercion, Two Axes, and the Importance of the Coercer. *Journal of Moral Philosophy* 5, 3, 394–422. http://doi.org/10.1163/174552408X369736

[5] Scott Anderson. 2010. The Enforcement Approach to Coercion. *Journal of Ethics and Social Philosophy* 5, 1.

[6] Mark B. Andrejevic. 2011. Surveillance and Alienation in the Online Economy. *Surveillance & Society* 8, 3, 278–287.

[7] Mark B. Andrejevic. 2011. Estrangement 2.0. *World Picture* 6, 1–14.

[8] Mark B. Andrejevic. 2012. Exploitation in the Data Mine. In *Internet and Surveillance*, Christian Fuchs, Marisol Sandoval, Boersma Kees and Anders Albrechtslund (eds.). Routledge, New York, N.Y, 71–88.

[9] Julia Angwin. 2014. Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance. Henry Holt & Co, New York, N.Y.

[10] Thomas Aquinas. 1920. *Summa Theologica*. Fathers of the English Dominican Province, London.

[11] Nathan Brooks. 2005. *Data Brokers: Background and Industry Overview*. Congressional Research Service, The Library of Congress, Washington, D.C.

[12] Danielle Keats Citron and Frank Pasquale. 2014. The Scored Society. *Washington Law Review* 89, 1, 1–33.

[13] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. W3C. Retrieved from http://www.w3.org/TR/P3P/

[14] Mary J. Culnan. 1993. "How Did They Get My Name?": An Exploratory Investigation of Consumer Attitudes toward Secondary Information Use. *MIS Quarterly* 17, 3, pp. 341–363.

[15] James Curran, Natalie Fenton, and Des Freedman. 2012. *Misunderstanding the Internet*. Routledge, London.

[16] John Deighton and Lenora Kornfield. 2012. *Economic Value of an Advertising-supported Internet Ecosystem*. Internet Advertising Bureau, New York, N.Y. Retrieved from http://www.iab.net/insights_research/industry_data_and_landscape/economicvalue

[17] Jenny van Doorn and JannyC. Hoekstra. 2013. Customization of Online Advertising: The Role of Intrusiveness. *Marketing Letters* 24, 4, 339–351. http://doi.org/10.1007/s11002-012-9222-1

[18] Tony Dyhouse. 2010. Addressing the Silo Mentality. *Infosecurity* 7, 2, 43. http://doi.org/10.1016/S1754-4548(10)70043-7

[19] Fredrik Erlandsson, Martin Boldt, and Henric Johnson. 2012. Privacy Threats Related to User Profiling in Online Social Networks. *Proceedings of 2012 International Conference on Social Computing*, IEEE, 838–842.

[20] Facebook Inc. 2014. *Facebook Inc. First Quarter 2014 Results*. Facebook Inc. Retrieved from http://investor.fb.com/releasedetail.cfm?ReleaseID=842071

[21] Federal Trade Commission. 2014. *Data Brokers: A Call for Transparency*. Federal Trade Commission.

[22] Jeremy Fogel. 2014. A Reasonable Expectation of Privacy. *Litigation* 40, 4. Retrieved December 15, 2014 from http://www.americanbar.org/publications/litigation_journal/2013-14/spring/a_reasonable_expectation_privacy.html

[23] Christian Fuchs. 2013. Class and Exploitation on the Internet. In *Digital labor - the Internet as playground and factory*, Trebor Scholz (ed.). Routledge, New York, 211–224.

[24] Christian Fuchs. 2014. *Digital Labour and Karl Marx*. Routledge, Taylor & Francis Group, New York, NY.

[25] Christian Fuchs and Marisol Sandoval. 2014. Digital Workers of the World Unite! A Framework for Critically Theorising and Analysing Digital Labour. *tripleC* 12, 2, 486–563.

[26] Ilana Gershon. 2011. Un-Friend My Heart: Facebook, Promiscuity, and Heartbreak in a Neoliberal Age. *Anthropological Quarterly* 84, 4, 865–894.

[27] James Jerome Gibson. 1986. *The Ecological Approach to Visual Perception*. Psychology Press, New York.

[28] Erving Goffman. 1990. *The Presentation of Self in Everyday Life*. Doubleday, New York, NY.

[29] Google Inc. 2014. *Google Inc. First Quarter 2014 Results*. Google Inc. Retrieved from http://investor.google.com/earnings/2014/Q1_google_earnings.html

[30] Martin Heidegger. 1977. The Question Concerning Technology, and Other Essays. Harper & Row, New York.

[31] M Hildebrandt. 2008. Defining Profiling: A New Type of Knowledge? In *Profiling the European Citizen*, M Hildebrandt and Serge Gurtwith (eds.). Springer, New York, NY.

[32] Simone van der Hof and Corien Prins. 2008. Personalisation and its Influence on Identities, Behaviour and Social Values. In *Profiling the European Citizen*, M Hildebrandt and Serge Gurtwith (eds.). Springer, New York, N.Y, 111–127.

[33] Gordon Hull, Heather Richter Lipford, and Celine Latulipe. 2011. Contextual Gaps: Privacy Issues on Facebook. *Ethics and Information Technology* 13, 4, 289–302. http://doi.org/10.1007/s10676-010-9224-8

[34] Ian Hutchby. 2001. Technologies, Texts and Affordances. *Sociology* 35, 2, 441–456. http://doi.org/10.1177/S0038038501000219

[35] Veikko Ikonen, Minni Kanerva, Panu Kouri, Bernd Stahl, and Kutoma Wakunuma. 2010. *D.1.2. Emerging Technologies Report*. ETICA Project.

[36] C. G. Jung. 1977. The Collected Works of C. G. Jung. 6: Psychological types. Pantheon Books, New York, NY.

[37] Paul A. Karger and Andrew J. Herbert. 1984. An Augmented Capability Architecture to Support Lattice Security and Traceability of Access. IEEE, 2–2. http://doi.org/10.1109/SP.1984.10001

[38] Helen Kennedy. 2008. New Media's Potential for Personalization. *Information, Communication & Society* 11, 3, 307–325. http://doi.org/10.1080/13691180802025293

[39] Byungwan Koh. 2011. *User Profiling in Online Marketplaces and Security*. ProQuest LLC, Ann Arbor, MI.

[40] Bert-Jaap Koops and Ronald Leenes. 2014. Privacy Regulation Cannot be Hardcoded. A Critical Comment on the "Privacy by Design" Provision in Data-Protection Law. *International Review of Law, Computers & Technology* 28, 2, 159–171. http://doi.org/10.1080/13600869.2013.801589

[41] Nicole C. Krämer and Stephan Winter. 2008. Impression Management 2.0: The Relationship of Self-Esteem, Extraversion, Self-Efficacy, and Self-Presentation Within Social Networking Sites. *Journal of Media Psychology* 20, 3, 106–116. http://doi.org/10.1027/1864-1105.20.3.106

[42] Hanna Krasnova, Oliver Günther, Sarah Spiekermann, and Ksenia Koroleva. 2009. Privacy Concerns and Identity in Online Social Networks. *Identity in the Information Society* 2, 1, 39–63. http://doi.org/10.1007/s12394-009-0019-1

[43] Steffan Krüger and Jacob Johanssen. 2014. Alienation and Digital Labour—A Depth-Hermeneutic Inquiry into Online Commodification and the Unconscious. *tripleC* 12, 2, 632–645.

[44] D. J. Lair. 2005. Marketization and the Recasting of the Professional Self: The Rhetoric and Ethics of Personal Branding. *Management Communication Quarterly* 18, 3, 307–343. http://doi.org/10.1177/0893318904270744

[45] Marc Langheinrich. 2001. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *Proceedings of the Third International Conference on Ubiquitous Computing*, Springer-Verlag, 273–291. Retrieved from http://www.vs.inf.ethz.ch/publ/papers/privacy-principles.pdf

[46] David Lazarus. 2015. Verizon's Super New Way to Mess with Your Privacy. *The Los Angeles times*.

[47] Yabing Liu, Krishna P Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing Facebook Privacy Settings: User Expectations vs. Reality. *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement*, ACM, 61–70.

[48] Michelle Madejski, Maritza Johnson, and Steven M Bellovin. 2012. A Study of Privacy Settings Errors in an Online Social Network. *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*, IEEE, 340–345.

[49] Alex Marthews and Catherine Tucker. 2014. Government Surveillance and Internet Search Behavior. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2412564

[50] Silva Martin and Stef Nicovich. 2013. Self Concept Clarity and its Impact on the Self-Avatar Relationship in a Mediated Environment. *Atlantic Marketing Association Conference Proceedings*.

[51] Viktor Mayer-Schönberger. 2009. *Delete: the Virtue of Forgetting in the Digital Age*. Princeton Univ. Press, Princeton, NJ.

[52] Damon McCoy, Kevin Bauer, Dirk Grunwald, Tadayoshi Kohno, and Douglas Sicker. 2008. Shining Light in Dark Places: Understanding the Tor network. *Privacy Enhancing Technologies*, Springer, 63–76.

[53] Soraya Mehdizadeh. 2010. Self-Presentation 2.0: Narcissism and Self-Esteem on Facebook. *Cyberpsychology, Behavior, and Social Networking* 13, 4, 357–364. http://doi.org/10.1089/cyber.2009.0257

[54] Edmund Mierzwinski and Jeffrey Chester. 2013. Selling Consumers Not Lists: The New World of Digital Decision-Making and the Role of the Fair Credit Reporting Act. *Suffolk University Law Review* 46, 3, 855–867.

[55] Hugh Miller and Jill Arnold. 2001. Self in Web Home Pages. In *Towards CyberPsychology*, Giuseppe Riva and Carlo Galimberti (eds.). IOS Press, Oxford, 73–94.

[56] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and Controlling the Filter Bubble through Interactive Visualization: a User Study. *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, ACM Press, 107–115. http://doi.org/10.1145/2631775.2631811

[57] Nicholas Negroponte. 1996. *Being Digital*. Vintage Books, New York, NY.

[58] Donald A. Norman. 2002. *The Design of Everyday Things*. Doubleday, New York.

[59] OECD. 2013. *Exploring the Economics of Personal Data*. Retrieved June 23, 2015 from http://www.oecd-ilibrary.org/science-and-technology/exploring-the-economics-of-personal-data_5k486qtxldmq-en

[60] Bertell Ollman. 2001. *Alienation: Marx's Concept of Man in Capitalist Society*. Cambridge University Press, Cambridge; New York.

[61] Rolf Oppliger (ed.). 2014. *Secure Messaging on the Internet*. Artech House, Boston.

[62] Karl Palmås. 2011. Panspectric Surveillance and the Contemporary Corporation. *Surveillance & Society* 8, 3, 338–354.

[63] Michael Rader, A. Antener, Rafael Capurro, Michael Nagenborg, and L. Stengel. 2010. *D.3.2. Evaluation Report*. ETICA Project.

[64] Joel Reidenberg. 2005. Technology & Internet Jurisdiction. *University of Pennsylvania Law Review* 153, 1951–1974.

[65] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. 2013. Bursting Your (Filter) Bubble: Strategies for Promoting Diverse Exposure. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, ACM Press, 95. http://doi.org/10.1145/2441955.2441981

[66] P.J. Rey. 2012. Alienation, Exploitation, and Social Media. *American Behavioral Scientist* 56, 4, 399–420. http://doi.org/10.1177/0002764211429367

[67] Jan-Willem van der Rijt. 2012. *The Importance of Assent: a Theory of Coercion and Dignity*. Springer, Dordrecht.

[68] George Ritzer. 2011. *Sociological Theory*. McGraw-Hill, New York.

[69] Ira S. Rubinstein. 2013. Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law* 3, 2, 74–87.

[70] Marisol Sandoval. 2012. A Critical Empirical Case Study of Consumer Surveillance on Web 2.0. In *Internet and Surveillance*, Christian Fuchs, Marisol Sandoval, Boersma Kees and Anders Albrechtslund (eds.). Routledge, New York, N.Y, 147–169.

[71] Kellyton dos Santos Brito, Frederico Araujo Durao, Vinicius Cardoso Garcia, and Silvio Romero de Lemos Meira. 2013. How People Care About Their Personal Data Released on Social Media. *Proceedings of 2013 Eleventh Annual International Conference on Privacy, Security and Trust (PST)*, IEEE, 111–118. http://doi.org/10.1109/PST.2013.6596044

[72] Amy J Schmitz. 2015. Secret Consumer Scores and Segmentations: Separating "Haves" from "Have-Nots." *Michigan State Law Review* 2014, 5, 1411.

[73] Bruce Schneier. 2015. *Data and Goliath: the Hidden Battles to Capture Your Data and Control Your World*. Norton, New York, NY.

[74] Richard C. Sherman. 2001. The Mind's Eye in Cyberspace: Online Perceptions of Self and Others. In *Towards CyberPsychology*, Giuseppe Riva and Carlo Galimberti (eds.). IOS Press, Oxford, 73–72.

[75] Daniel J. Solove. 2004. *The Digital Person: Technology and Privacy in the Information Age*. New York Univ. Press, New York, NY [u.a.].

[76] Inger L. Stole. 2014. Persistent Pursuit of Personal Information: A Historical Perspective on Digital Advertising Strategies. *Critical Studies in Media Communication* 31, 2, 129–133. http://doi.org/10.1080/15295036.2014.921319

[77] John Suler. 2004. The Online Disinhibition Effect. *CyberPsychology & Behavior* 7, 3, 321–326. http://doi.org/10.1089/1094931041291295

[78] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2208240

[79] Omar Tene and Jules Polonetsky. 2012. To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising. *Minnesota Journal of Law, Science & Technology* 13, 1, 281–358.

[80] Omer Tene and Jules Polonetsky. 2013. Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* 11, 5, 239–272.

[81] Joseph Turow. 2011. *The Daily You: How the New Advertising Industry is Defining Your Identity and Your World.* Yale University Press, New Haven.

[82] Shaojung Sharon Wang. 2013. "I Share, Therefore I Am": Personality Traits, Life Satisfaction, and Facebook Check-Ins. *Cyberpsychology, Behavior, and Social Networking* 16, 12, 870–877. http://doi.org/10.1089/cyber.2012.0395

[83] Logan Danielle Wayne. 2012. The Data-Broker Threat. *Journal of Criminal Law and Criminology* 102, 1, 253–282.

[84] Alan F Westin. 1970. *Privacy and Freedom*. Bodley Head, London.

[85] S. Wilske and T. Schiller. 1997. International Jurisdiction in Cyberspace: Which States may Regulate the Internet? *Federal Communications Law Journal* 50, 119 – 178.

[86] Allen W. Wood. 2006. *Karl Marx*. Routledge, New York.

[87] Mike Z. Yao and Daniel G. Linz. 2008. Predicting Self-Protections of Online Privacy. *CyberPsychology & Behavior* 11, 5, 615–617. http://doi.org/10.1089/cpb.2007.0208

[88] G. I. Zekos. 2006. State Cyberspace Jurisdiction and Personal Cyberspace Jurisdiction. *International Journal of Law and Information Technology* 15, 1, 1–37. http://doi.org/10.1093/ijlit/eai029

[89] Shanyang Zhao, Sherri Grasmuck, and Jason Martin. 2008. Identity construction on Facebook: Digital Empowerment in Anchored Relationships. *Computers in Human Behavior* 24, 5, 1816–1836. http://doi.org/10.1016/j.chb.2008.02.012

Era of Big Data: Danger of Descrimination

Andra Gumbus
Professor Management
Sacred Heart University
5151 Park Ave
Fairfield, CT 06825
gumbusa@sacredheart.edu

Frances Grodzinsky
Professor Computer Science
Sacred Heart University
5151 Park Ave
Fairfield, CT 06825
grodzinskyf@sacredheart.edu

ABSTRACT

We live in a world of data collection where organizations and marketers know our income, our credit rating and history, our love life, race, ethnicity, religion, interests, travel history and plans, hobbies, health concerns, spending habits and millions of other data points about our private lives. This data, mined for our behaviors, habits, likes and dislikes, is referred to as the "creep factor" of big data [1]. It is estimated that data generated worldwide will be 1.3 zettabytes (ZB) by 2016. The rise of computational power plus cheaper and faster devices to capture, collect, store and process data, translates into the "datafication" of society [4]. This paper will examine a side effect of datafication: discrimination.

Categories and Subject Descriptors
K.4.1. **[Computers and Society**]: Ethics

General Terms
Human Factors

Keywords
Big Data, Discrimination, Human Resources, Privacy

INTRODUCTION

We live in a world of data collection where organizations and marketers know our income, our credit rating and history, our love life, race, ethnicity, religion, interests, travel history and plans, hobbies, health concerns, spending habits and millions of other data points about our private lives. This data, mined for our behaviors, habits, likes and dislikes, is referred to as the "creep factor" of big data [1].

It is estimated that data generated worldwide will be 1.3 zettabytes (ZB= 2^70) by 2016. The rise of computational power plus cheaper and faster devices to capture, collect, store and process data, translates into the "datafication" of society [4].

This paper will examine a side effect of datafication: discrimination. The first part will analyze discriminatory practices based on profiling. Next, it will relate privacy concerns to discriminatory practices, and finally, it will examine the impact of Big Data on Human Resource departments within organizations.

According to the National Institute of Standards and Technology Big Data is data which: "exceed(s) the capacity or capability of current or conventional methods and systems" [2]. A proponent of Big Data, Alex Pentland, Director of the Media Lab Entrepreneurship Program at MIT, describes Big Data as a new asset that we are just beginning to understand. He believes that it is a quantitative measure of human behavior that can be effectively used to solve human problems [2]. Other proponents think that modern economic activity is dependent on Big Data for the functioning of our global economy. Bringing together pools of data to analyze patterns and make informed decisions is the basis for competition and growth as well as enhanced productivity and value creation in business.

Data from industrial goods are being analyzed to provide better service and design of products based on actual use. "The ability to "now cast" using real time data enables prediction and theory testing never before possible in applications in the public sector and in personal location data" [3]. While we acknowledge that developments in the use of Big Data may have the capacity to promote social good, we claim that they also can also perpetuate harm with results that are inequitable or discriminatory when applied to protected classes. Big data analytics can lead to outcomes that go against civil liberties like fair housing, employment, credit and consumer protection.

In their book *Big Data,* Mayer-Schonberger and Cukier describe this new age of big data on page 97.

> Today we are a numerate society because we presume that the world is understandable with numbers and math, and we take for granted that knowledge can be transmitted across time and space. Future generations may have a big data consciousness and the presumption that there will be a quantitative component to everything. … in the new age of data, all data will be regarded as valuable [4].

Sectors such as online advertising, health care utilities, transport, logistics and public administration are using big data to stimulate innovation and productivity growth. Data driven R&D provides enhanced research and development; data-intensive product development uses data as a product or as a component of a product; data–driven processes can optimize production or delivery processes; data-driven marketing improves efforts by targeting ads and personalizing recommendations; and finally data is used to improve management practices and approaches [5].

1. BIG DATA SOCIETY: DISCRIMINATORY PRACTICES

Zwitter identified three categories of Big Data stakeholders. First are the collectors who determine what is collected and how long it is kept; next the utilizers who define and redefine the purpose for use of the data, and finally those of us who generate data. He defines data generators as those who input or record data voluntarily or unknowingly [6]. Data generators are at a disadvantage by not knowing who is collecting data about them and by not knowing how that data is being used. Power inequality exists between the generator and the collector and utilizer, both of whom have greater power than the generator. The Internet of Things (IoT) and global data exacerbates the power imbalance benefitting corporate entities who know how to generate intelligence from data [6].

Under civil rights law, discrimination can occur when there is disparate treatment, with disparate impact. Disparate treatment results from treating a person differently on the basis of race, gender, age, religion or other protected classes. Disparate impact results from a policy or practice that has a disproportionate negative effect on a protected class [7]. Existing anti-discrimination laws in the United States prohibit use of data that will discriminate based on health or disability. For example, employers cannot legally refuse to hire or fire someone who has an illness. However, there is nothing to stop employers with access to data from determining the probability of illness or disease based on health and eating habits. These employees could then be viewed as expensive, a potential insurance risk and therefore non-desirable. How is this done in company practice?

> The analyst involved, whether inside or outside the firm could easily mask the use of health-predictive information. A firm could conclude a worker is likely to be diabetic and a "high cost worker" given the cost of medical care. Given the proprietary nature of the information involved, the most the firm will tell the un-hired or fired worker is the end result: the data predicted that cost to the firm was greater than value (if a rationale is offered)...Secrecy is a discriminator's best friend: unknown unfairness can never be challenged, let alone corrected (page 1421) [8].

Applications of Big Data are designed to differentiate between different types of people and make distinctions that separate desirable from undesirable individuals when it comes to credit risk, mortgage awards, credit card issuance or customer pricing. The mining of behavioral data carries the risk of the statistical problem of false positives when individuals are placed in a group that grants them undeserved privileges or false negatives: when individuals are placed in a category that inadvertently harms them. When this occurs some are disadvantaged and some have an unfair advantage despite the assertion that data mining algorithms have a 99% accuracy rating. As a result, the resulting misinterpretations may constitute wrong treatment for hundreds of thousands of people who might fall in that 1%.

Much of the problem of discriminatory practice has to do with how the results of Big Data analysis are interpreted and used. The sheer quantity of produced data has given rise to an industry of companies that will help you make sense of analysis results. Those who encourage us to believe that correlations are infallible may be ignoring the fact that their use in particular contexts may be dangerous. For example, some results may give rise to the possibility of profiling based on age, race, sexual orientation or other characteristics and behaviors which when correlated could lead to discriminatory practices. As a result, disparate impact or unequal treatment of an identified class compared to similar groups could result from data analytics. Murphy [9] reveals that job applicants are being profiled using references, prior employment, credit rating, driving record, criminal record credit history, Facebook pages and other sources that can impact hiring decisions that breach employment laws [9]. When correlations lead to policy based on profiled categories the possibility for discrimination exists. Nathan Newman believes,

> Economic inequality is driven by inappropriate use of big data which can coincide with the economic downturn and loss of income for average households. There are other factors contributing to inequality such as de-unionization, globalization and the automaton of unskilled jobs, but when combined with data consolidation the harm to low income and other vulnerable segments of the population increases [10].

His view is supported by FTC Chairwoman Edith Ramirez who stated that big data has the capacity to reinforce disadvantages faced by low income and underserved communities and called for greater transparency and accountability to make sure that low income populations would not receive differential treatment through digital redlining and discrimination by algorithm. Existing disparities can be exacerbated by the segmentation of customers to determine what products are marketed to them, what prices are quoted and what level of service they receive. Conscientious policy makers should ensure that Big Data be used for economic inclusion, not exclusion [11].

Big Data platforms enable racial profiling in subtle and invisible ways by targeting home address and other characteristics as a proxy for race. Online discrimination steered approximately 30,000 Black and Hispanic lenders into costly subprime mortgages during 2004 – 2009 and charged them higher fees than white lenders [10]. These targeted customers were disproportionally Black and Latino and were offered mortgages that had 30% higher interest rates compared to White borrowers. Unethical companies can target vulnerable less educated populations to mislead them with scams of harmful offers. The data industry uses the term "sucker lists" or "suffering seniors" who have been identified as targets for unethical and misleading scams. Algorithmic profiling allows companies to discriminate and categorize consumers into profiled groups in ways that may

harm them with price discrimination and other unwelcome exploitive marketing practices [10]. Major corporations such as Staples, Home Depot and other financial services organizations use user location to display different prices to different customers. Instead of benefitting the low income population with lower prices, they did the reverse charging low income people higher prices and giving higher income people better deals. Credit card companies have similar practices offering different deals based on locations and presumptions about income levels. When retailers obscure prices and discriminate, economic models show that prices are higher than if consumers knew all the prices [10]. Price obfuscation strategies foster economic inequality and harm the least well off.

The asymmetry of power between data mining companies and individuals results in a data imbalance between the data have's (government and large corporations) and have not's. Perfect personalization or profiling can result in policies that discriminate for products and services or pricing of products. Buchta [1] explains that this gives companies the potential to create a perfect bubble for each consumer, presenting him/her with only information that algorithms dictate are of value. A cost benefit analysis falls short on the benefits when the harms are factored into this equation. She calls for greater regulation of the data gathering industry, more transparency, notice and choice for consumers [1]. Pasquale and Citron note (page 1419) that

> Of great concern is the collection and analysis of a critical mass of data. Our lives are starting to become an open book for those powerful or rich enough to score our profiles…Will individuals hesitate to join mental health support groups… will they refrain from joining political groups once they realize their affiliations on social media are a detriment to their careers? [8]

Posts on social network sites, locations from smartphones, sensors in our homes and on our bodies create a "nearly ubiquitous data collection capability that can erode our civil liberties and foster discrimination" [12]. Google searches for people with African-American sounding names were more likely to display ads with the word "arrest" which could lead to unfair and inaccurate perceptions of the person. The Chicago police department mined social networks and found 400 people who a model deemed likely to be involved in violent crime. Innocent people run a greater risk of being profiled by computer algorithms [12].

As our data is collected, interpreted and used without our consent, questions about fairness arise. What actually affects our lives in society? To make this point, Helbing asks the following:

- …How can you be sure you are getting your loan for fair conditions, and do not pay a higher interest rate because someone in your neighborhood defaulted?
- Can you afford to live in a multicultural quarter or should you move to a neighborhood to get a reasonable loan?
- Is there a tariff on your health insurance or do you pay more because your neighbors do not jog?
- Should you drink that extra glass of wine, eat red meat or will your mortgage rate go up?
- Would there be a right way of living or would everyone be discriminated against for some behavior or get rewards for other behaviors? [13]

The answers to these questions are elusive. We do not know how much information is collected about us, how long it is kept and how it is used. At present, users have no control over what is collected about them, and this makes it difficult, therefore, to judge whether we have been victims of discriminatory practices. The consumer Watchdog writes that "…consumers deserve clear understandable standards for use of their information" [14].

2. PRIVACY ISSUES

The advent of the IoT means that virtually anything connected to the Internet (TV, phone, tablet, refrigerator, camera, and car) provides data in the IoT movement [15]. Baker identified four major shifts in data collection that erode privacy. They are: invasiveness, variety, integration and scope. Government and businesses collect increasing amounts of data irrespective of privacy boundaries. Data sources are expanding as social media and machine data proliferate. Data is gathered for knowledge's sake not just under the guise of better customer service, marketing, or security. With privacy regulation much of the data is de-identified stripped of name and address or other identifying markers. The problem lies in the re-identification which is very easy to perform using mobile device ID and IP addresses. Data gathered with an IP address can predict a zip code which can be used as a proxy for race and income. The concept of personally identifying information such as social security number and credit card numbers is changing now that we can directly identify individuals based on the volume of data they generate. Computer scientists at Carnegie Mellon predicted full nine digit social security numbers for 8.5 % of people born in the US between 1989 and 2003 [16].

Under Fair Information Practice Principles (FIPP) privacy policy, companies must give consumers notice of data they collect, why they collect it and who they will share it with. They are supposed to use the data only for the purposes for which it was collected and not secondary uses. The Center for Democracy and Technology states that privacy laws should empower people to make informed choices about how their data is collected, stored, used, shared and maintained. The Computer & Communications Industry Association recommends a balance between the benefits and concerns of Big Data. It believes the focus should be on harms that occur from misuse and implications from who is using data, under what terms and for what purpose. Public interest groups also call for special protection for sensitive categories such as financial information, health, race, ethnicity, geo-location, age and data collected in the educational context [17].

How do we protect ourselves from the arbitrariness that can result from informational injustice when data is mined inaccurately? One approach is to legislate or establish a government agency with standards and certification procedures or punishments for violation to guard against false conclusions from data mining. Another is to equip individuals with the ability to correct data or run their own scenarios using various algorithms to run simulations in order to see what predictions result. Helbing describes this as a transparent and participatory approach where results can be verified or falsified, enabling trust in the

algorithms and enhanced quality of healthy data results. Citizens control their data and participate in the value generated by their data. They can comment, correct, and determine what kinds of data are used for what purpose enhancing privacy and self-determination [13].

3. WHEN BIG DATA MEETS HUMAN RESOURCES

The human resource function is responsible for guaranteeing that the organization does not discriminate in employment practices and for making sure that state and federal laws as well as company policies are enforced. Overall however, legal and ethical issues are not widely discussed in the research on HR data mining; however, topics of privacy and equality can be found in the literature. Avoiding discrimination and treating people equitably means avoiding unfair treatment based on membership in a group. Discrimination can be blatant or hidden. Stereotyping is a problem in mining data to make HR decisions when unfair and unequal treatment is based on algorithms assigning classification and segmenting individuals into groups based on data. This often occurs without the knowledge of the owner of the data.

What is the role of data in making human resource decisions and what possible discrimination can result from the use of data when recruiting, selecting and making employment decisions? Anything but raw data may not be free from human bias. Human bias could affect what data is collected, what variables are included, what sources are used, what is mined versus what is ignored, and what questions are prioritized. Cold data could also be polluted and corrupted by ingrained company practices or design of algorithms [15]. How do we make sense of all this data? The new companies, whose sole purpose is to help organizations realize business value from their data, do not usually address the ethical issues.

Human resource departments assemble data on factors such as employee attrition and hiring, compensation and benefits, ethnic, gender, cultural, and nationality distributions. By applying advanced analytical techniques on the data, human resource professionals can get business insight, predict changes, and make informed decisions at operational and strategic levels [18]. Online analytical processing and data mining focus on past performance; predictive analytics forecasts on future behavior in order to guide decisions. Data mining tells us what has happened while predictive analytics advise us on appropriate response action. Key activities such as trends, metrics, and performance indices are portrayed in scorecards and dashboards. Advanced analytics can answer human resource questions such as whether capital investments contribute to business performance, how much human resource activities impact employee performance, or what skills the organization will need to meet future opportunities [18].

Big Data has entered the field of human resource management where analysis of the data guides the hiring, promotion and career planning functions in a new field called "work-force science". This is done through the analysis of email, instant messaging, phone calls, written code and mouse clicks, mined to determine how people work and, who they are connected to in their social network. Personality based assessments and other tools and tests used in selection and hiring decisions can be aggregated to determine worker communication patterns, style, and results. The proponents of work-force science predict that it will lead to efficiency and innovation within companies that traditionally rely on gut feel, interviews and reference checking to make hiring decisions. They believe that the revolution in measurement resulting from Big Data will change organizational and personnel economics. They predict that work-force science will "be applied across the spectrum of jobs and professions, building profits, productivity, innovation and worker satisfaction" [19]. However, worker surveillance raises many questions of employee privacy, ownership of data and the use and interpretation of that data. One ethical problem is that usually there is no informed consent about collection and use of this data even though it is being used to make important career decisions that impact worker livelihood.

In order to search for top talent, human resources go to analytics firms that assess talent and provide scores of a candidate in various fields. For example, a candidate's online contributions can be tracked by Remarkable Hire that provides a hiring score or Talent Bin and Guild that provides lists of potential applicants based on online data [20]. HR departments are using computer games and tests to measure emotional intelligence, memory, creativity, knowledge and cognition and employees' willingness to take risks. Companies like Google who previously used SAT's and GPA scores found that these did not correlate to success at Google [19]. They are now using additional metrics. For example, for a programming job, recruiters looked at how well the person codes; is the code reusable and is it respected among other programmers? Companies are now mainly using work-force science in call centers to analyze hourly workers in order to reduce attrition rates which are common at 100%. In these types of settings the improvement opportunity and cost savings is great. With the cost of hiring averaging $1500 per hire, a company found it could hire 800 instead of 1000 people and still had 500 workers on the job 3 months later. It claimed better customer service and less worker-churn [19].

3.1. Dangers of Big Data in Human Resources

In the area of training and development, Big Data can be used to benefit companies in areas such as: the identification of who might leave the organization; retention of top talent; the ability to identify top potentials for succession planning; the ability to assess what drives performance. Based on these metrics, they can adjust their management style. However, a simple misuse or mistake regarding reward or promotion based on an algorithm can have serious negative consequences for the organization as well as the employees if data is mishandled.

Race, gender, ethnicity, age and other discriminatory hiring practices have plagued HR in the past. Proponents of Big Data analytics advise that the crunching of thousands of bits of data may help to eliminate bias by offering 300 variables giving us a more robust portrait of the candidate. Because of the volume of available data, traditional screens like college attended, recommendations from fellow employees or previous employers can be combined with new screens such as "the sites where a person hangs out, the types of language used to describe technology, and self-reported skills on LinkedIn, projects worked on [19]. Some recruiters are using communication styles as a significant metric: What is the person's communication pattern?

How does he/she present on social media sites, and how does he/she communicate ideas?

We are not sure that this is all good news. The recruiters who use social media sites for data can gain disturbing insights from non-work related sites. For example, a student of ours was denied an internship based on old high school photos posted on Facebook that he had neglected to remove. The practice of using non-traditional screens in HR has resulted in law suits from victims who feel they were denied an opportunity and discriminated against in the employment process. There are some protections in U.S. law to protect potential job candidates: the Human rights Act 1998 provides a respect for private and family life; the Data protection Act 1998 states that data holders not have excessive information nor process it unfairly; the Civil Rights Act of 1964 and 1991 protects discrimination by gender, race, national origin, sexual orientation; the Age Discrimination in Employment Act (ADEA) protects against age discrimination and the Americans with Disabilities Act (ADA) against unfair treatment because of disability.

Besides the existing laws, how can we insure ethical use of Big Data in HR practices involving employment decisions? Kettleborough recommends four considerations. First, quality and accuracy must be assured when making life changing decisions about employees and candidates. Second, there must be enough data to make informed decisions and understand probability, sample size and statistical significance. Third, there must be caution about correlation and causation conclusions, i.e., two items that correlate do not necessarily cause each other. Finally, privacy and anonymity must be safe guarded so that personal data is not used against individuals. Also for internally administered surveys on employee satisfaction and culture, we must guard against using demographic information to identify individuals in a way that might turn honest data into dangerous data [21].

Peoplefluent produced a white paper outlining how HR departments can unlock data's value and be more proactive in preparing their organizations for the era of Big Data. Because of the large number of sources, data integration can be a problem. Those companies who have found effective ways to integrate their data have shown more success [22]. Human Resources can jumpstart data mining efforts and be a role model for other functions in an organization. They recommend using a role-based approach to analyzing people data based on functional roles in HR using the following six roles: compensation manager, chief learning officer, line of business manager, and VP HR/ head of talent management [23]. Compensation managers analyze reward schemes and compensation programs in order to ensure accuracy, fairness and visibility to employees. Learning officers look at training needs and data to ensure that employees have the right tools and raining at the right intervals to perform their jobs. The recruitment function looks at identifying optimal internal and external candidates to accelerate the hiring process. The procurement officer projects contingent workforce needs and look at staffing requirements and sourcing resources. Business managers are concerned with managing performance against company goals. The head of HR and talent management is responsible for data from all functional areas to determine if HR is hitting goals and contributing to organizational success. Using predictive analysis to assess historical data and influence future outcomes can enable HR to drive results strategically and be proactive partners in the business as long as they take measures to avoid discriminatory practices.

4. USING BIG DATA WITH HUMILITY AND HUMANITY

In May, 2014 the White House issued a report recommending government limits on how companies make use of information they gather from online customers. The report makes six policy recommendations including a national data breach law that requires disclosure when personal credit card data is exposed and defines customer rights regarding how their data is used. This protection extends to non-citizens of the US and to students regarding educational data [24]. An important aspect of the report is the acknowledgement that data misuse can be discriminatory. Misuse of data has " The potential to eclipse longstanding civil rights to protections on how personal information is used in housing, credit, employment, health, education and the marketplace" [24]. Assessing human values and recognizing the limitations of Big Data are critical for its ethical use.

Mayer-Schonberger and Cukier [4] predict that the effect on individuals may be the most harmful aspect of our future reliance on Big Data. They caution us that individual expertise matters less in a world where probability and correlation are paramount. "The danger to us as individuals shifts from privacy to probability. This leads to an ethical consideration of the role of free will versus the dictatorship of data. We will need new rules to safeguard the sanctity of the individual" (page 17). The authors warn that the demarcation between measurement and manipulation is blurred by the vast amount of data collected, and by our inability to conceptualize just what constitutes Big Data or how it is being used. Technology has reached a point where vast amounts of information can be captured and recorded cheaply. Data can be collected passively, and because the cost of storage has fallen, it is easier to justify keeping this data. Over the past half century the cost of digital storage has roughly been reduced by half every two years while storage density has increased 50 million fold [4].

In order to combat the dominance of Big Data gathering companies, consumers need more control of their data and possible government interventions to protect them. Strategies used in the past to protect consumers such as notice and consent, opting out, and anonymization are no longer effective based on the volume of data available. Users are easily identified and advertisers can fingerprint Web browser according to their skills. Individuals can be re-identified from anonymous data using zip code, birth date and gender to an 87.1% accuracy [9].

These problems can also be addressed by empowering individuals with access to their data and allowing them to analyze their own data and make conclusions from it. This sharing the wealth strategy can address Big Data privacy concerns by empowering consumers and represents a shift in the business model from organizations owning data to individual control. Consumers become free and independent actors in the marketplace, telling vendors what they want; how they want it, when and at what price [25]. This consumer centric model gives individuals control over management and use of their data, selective disclosure of selective data, control over purpose and duration of use, and correlations permitted by the individual not the end user. It also provides for a high level of security, data portability and accountability and enforcement. The question remains whether we can address challenges of this new business

model such as technical feasibility, intellectual property rights, and business incentives to switch to a new paradigm [25].

One strategy was tried by Acxiom, the largest data broker, in 2013. Acxiom let people see what information it had about them in a Web site AboutTheData.com. When accessed, the site revealed core data Acxiom had amassed in an effort toward transparency by data brokers. Critics claim that Acxiom revealed selective facts only and not the analysis the company markets to clients such as categories like "potential inheritor," "adult with senior parent," and "diabetic focus" [25].

Another strategy to introduce humility and humanity into the equation is to ensure algorithmic accountability by having closer human scrutiny of the results of algorithms used to make life - changing decisions. Big Data is supposed to bring greater economic opportunity and convenience to all people not just a preferred few. With human oversight adding "machine-to-man" translation of results, data equality will become a reality. It will give context to analytic results. Predictive recommendations can be reviewed and overruled in essence giving human veto power over the result. Critics of data science may object to human intervention, yet this introduces an element of protection for the individual (page A4) [26].

> In a sense a math model is the equivalent of a metaphor, a descriptive simplification. It usefully distills, but it also somewhat distorts. So at times, a human helper can provide that dose of nuanced data that escapes the algorithmic automation. Often the two can be better than the algorithm alone [26].

Gary King, Director of Harvard's Institute for Qualitative Social science recommends that the creators of the algorithms make adjustments in the design of the calculations to favor the individual in order to reduce the risk of getting a wrong result. It will also improve trust in predictive results if the process were more transparent (page A3).

> The key that will make it work and make it acceptable to society is storytelling. Not so much literal storytelling, but an understandable audit trail that explains how an automated decision was made. How does it relate to us? How much of this decision is the machine and how much is human? [26]

In sharp contrast to Big Data is Open Data which is accessible to everyone. Gurin defines Open Data as available to people, companies, and organizations that can be used to make data driven decisions and solve complex problems. The Open data model includes over 500 companies across business sectors that provide platforms to make government data easier to find and access [26]. Open Data is currently being used in legal services including patent data and competitive intelligence; education including data on value of institutions; energy efficiency; precision agriculture; health care transformation; housing and real estate and transportation analysis. The Open Data 500 study includes companies that earn revenue from a variety of business models serving diverse customers. As the amount of federal, state and local data increases the business opportunities will expand for data that is accessible to everyone. The goal of Open Data is to make all government data open unless privacy or security dictates otherwise [27].

CONCLUSION

Organizations that use Big Data analytics should practice it with customer privacy and integrity of data in mind, and guarantee legal and ethical applications through their policies and procedures on the use of data.

In the eSociety where everything has a score, predictive algorithms determine who has value and will receive critical life changing opportunities determined by score. Without fair and accurate scoring systems data can be biased and arbitrarily assign individuals to a stigmatizing group that affects their opportunities. Advances in artificial intelligence are missing the human element,, and we believe that human values are needed as oversight in the design and execution of scoring systems. We need to consider the consequences when we rely solely on scoring machines to make decisions that may not be fair or just.

Citron studied the scored society using credit score as a case study and found three basic problems with credit scores: opacity or lack of transparency, arbitrary results and disparate impact on women and minorities. Consumers do not know why or how their credit scores change. Different credit bureaus have vastly different scores for the same individual and punish cardholders for paying bills. Biases are embedded in the code and defined parameters of data mining. For example certain occupations can get a low score like service jobs which are held by minorities. Although discrimination was not intended, and may be unintentional, it is discrimination none the less. Credit scores have a negative disparate impact on disadvantaged groups – women and minorities as recent settlements by Allstate typify where five million African-American and Hispanic customers were discriminated against in the denial of insurance based on credit score [28].

Citron recommends regulatory oversight of scoring systems to include: gathering of data into scores, calculating gathered data into scores, disseminating scores to decision makers, and employers and others use of scores in making decisions. Ideally calculations would be public and processes transparent, inspected for fairness and accuracy. Individuals deserve to know how they are rated and who is getting the data. Licensing and audit requirements for sensitive areas that impact employment, insurance or heath care are needed to avoid arbitrariness by algorithm [28]. To this end the FTC addressed the following concerns about predictive algorithms: How are companies using scores? Are they accurate? Can consumers benefit from available scores? How is privacy ensured? Patterns and correlations about race, nationality, sexual orientation and gender that are already covered by discrimination law deserve added scrutiny

FTC Chairwoman Ramirez stated that decisions by algorithm require

> transparency, meaningful oversight and procedures to remediate decisions that adversely affect individuals who have been wrongly categorized by correlation. Companies must be sure that they are not using big data algorithms that are accidently classifying people based on categories that society has decided by law or ethics

not to use such as race, ethnic background, gender and sexual orientation [28].

As we utilize insights gained from Big Data analytics we need to recognize that results have a scope limited by context. Much of the data we generate is collected without a question in mind although it is being used to make predictions about us. Although correlations can be very useful, when it comes to interpreting them and making decisions, we are not willing to give over final decisions affecting individuals in society to a machine alone.

We need to recognize the perils of Big Data when decisions are made about disadvantaged and protected classes. We need to guard against data that reinforces gaps between the rich and poor, haves and have not's and that suppress already disadvantaged people and benefit the wealthy and privileged. We cannot succumb to the powerful allure of data only as precise and reliable, when it can also be unjust and unfair, constraining opportunities for the disadvantaged and perpetuating discrimination. The exponential growth of data has the capacity to bring great value to society but can challenge the ethical and legal systems if the rights of individuals are violated in the process of bringing added value to business.

REFERENCES

[1] Buchta, Heather (2014) How Did Data Get to Be So Big? Inside Counsel. Breaking News. November 25, http://www.insidecounsel.com/2014/11/25/how-did-data-get-to-be-so-big Accessed 6/1/15.

[2]MIT Technology Review (2013), http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/ , Accessed October, 3, 2013.

[3]McGuire,T., Manyika, J. and Chui, M. (2012)"Why Big Data is the New Competitive Advantage". Ivey Business Journal, Jul/Aug, Vol.76 Issue 4, pp. 1-4.

[4]Mayer-Schonberger, V. and Cukier, K. Big Data: A revolution that will transform how we live,work, and think. (2013) Houghton Mifflin Harcourt, Boston, NY.

[5]OECD Digital Economy Papers 222 (2013) Exploring Data-Driven Innovation as a New Sources of Growth: Mapping the Policy Issues Raised by Big Data, April.

[6]Zwitter, Andrej (2014) Big Data Ethics. Big Data & Society July – December, 2014: 1 – 6.

[7]Yu, Persis, McLaughlin,Jillian and Levy, Marina. (2014) Big Data: A big Disappointment for Scoring Consumer Credit Risk. National Consumer Law Center March 2014.

[8]Pasquale, Frank and Citron, Danielee Keats (2014) Promoting Innovation While Preventing Discrimination: Policy Goals for the Scored Society . Washington Law Review 89:1413.

[9]Murphy, Michael and Barton, Jophn. (2014) From a Sea of Data to Actionable Insights: Big Data and What it Means for Lawyers. Intellectual property & Technology Law Journal March, 26.3: 8 – 17.

[10]Newman, Nathan. (2014) How Big Data Enables Economic Harm to Consumers, Especially Low Income and Other Vulnerable Sectors of the Population. Journal of Internet law December, 18.6: 11 – 23.

[11]Curran, John (2014) FTC Chief Sounds Note of Caution on Development of Big Data. Cybersecurity Policy Report, September.

[12]Dwoskin, Elizabeth (2014) White House Takes Aim at Big Data Discrimination; Report recommends More Privacy laws. Wall Street Journal (Online) . May 1, 2014.

[13]Helbing, Dirk. (2014) Big Data Society: Age of Reputation or Age of Discrimination? http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2501356 Accessed May 1, 2014.

[14]Wireless News (2014) Consumer Watchdog Supports 6 Policy Recommendations in White House Big Data Report. May 7.

[15]Baker, Pam. (2015) Data Divination: Big Data Strategies. Cengage Learning PTR, Boston, MA

[16]Lohr, Steve (2015) Maintaining A Human Touch As the Algorithms Get to Work. New York Times, April 7, p. A3

[17]Hammond, Brian. (2014) Industry Groups Stress Need to Protect Innovation in Big Data privacy Effort. Telecommunications Reports, Sept 1,2014 80.17: 29 -32

[18]Kapoor, Bhushan.(2011), Impact of Globilization on Human Resource Management. Journal of International Management Studies 6.1 (Feb) : 1 – 8.

[19]Lohr, S. (2013). Big Data, Trying to Build Better Workers. New York Times, April 21, p. 5.

[20]Zarsky, Tal Z. (2014) Understanding Discrimination in the Scored Society, Washington Law Review, 89:1375.

[21]Kettleborough, Jonathan. (2014), Big Data. Training Journal. June. 14 – 19.

[22] Grossman, K (2014) "System-integration drives talent acquisition". http://www.peoplefluent.com/blog/hr-system-integration-drives-talent-acquisition, Accessed June 12, 2015.

[23]www.peoplefluent.com (2014) Make Your HR Data Actionable Now! Unlock the Value Trapped in Your Company's Data by using Role - Based Analytics. A Peoplefluent White Paper. Accessed May 14, 2014.

[24]Sanger, D. and Lohr, S. (2014). Call for Limits on Web Data of Customers. NYT May 2, 2014. P. A1 and B6.

[25]Rubinstein, Ira S. (2013) Big Data: The End of Privacy or a New Beginning? International Data privacy Law, Vol 3, No.2: 74 - 87

[26]Lohr, Steve (2015) Dataism: The Revolution Transforming Decision Making, Consumer Behavior, and Almost Everything Else. HarperCollins, NY,NY.

[27]Gurin, Joel (2015) Opening Business Innovation With Open Data. Business Horizon Quarterly. Issue 12, pp. 42 – 49.

[28]Citron,Danielle Keats and Pasquale, Frank (2014) The Scored Society: Due Process for Automated Predictions. March, Washington Law Review. 1 – 33.

.

Augmented Reality All Around Us: Power and Perception at a Crossroads

Marty J. Wolf
Bemidji State University
Bemidji, MN USA 56601
mjwolf@bemidjistate.edu

Frances Grodzinsky
Sacred Heart University
Fairfield, CT USA 06825
grodzinskyf@sacredheart.edu

Keith Miller
University of Missouri - St. Louis
St. Louis, MO USA 563121
millerkei@umsl.edu

ABSTRACT
In this paper we continue to explore the ethics and social impact of augmented visual field devices (AVFDs). Recently, Microsoft announced the pending release of HoloLens, and Magic Leap filed a patent application for technology that will project light directly onto the wearer's retina. Here we explore the notion of deception in relation to the impact these devices have on developers, users, and non-users as they interact via these devices. These sorts of interactions raise questions regarding autonomy and suggest a strong need for informed consent protocols. We identify issues of ownership that arise due to the blending of physical and virtual space and important ways that these devices impact trust. Finally, we explore how these devices impact individual identity and thus raise the question of ownership of the space between an object and someone's eyes. We conclude that developers ought to take time to design and implement a natural and easy to use informed consent system with these devices.

Categories and Subject Descriptors
K.4.1 **[Computers and Society]**: Ethics.

General Terms
Human Factors.

Keywords
Augmented Reality, Augmented Visual Field Devices, Autonomy, Deception, Human Values, Identity, Informed Consent, Trust.

1. FRAMING THE DISCUSSION
This paper extends and elaborates an earlier paper, Grodzinsky, miller and Wolf [1]. In that paper, we explored augmented visual field devices (AVFDs), using the following definition for *visual augmented reality* (AR): "…visual augmented reality involves projecting light in such a way that both natural light and artificial light enter the eye simultaneously, so that some objects seen in the visual field can be traced back to physical objects, and other objects seen are virtual objects, for which no physical object is the source of reflected light." Since that time, Microsoft announced the pending release of HoloLens, "the world's most advanced holographic computing platform" [2]. HoloLens seemingly will project holographic images into the physical space that are visible to at least the wearer of the HoloLens. In addition Magic Leap has recently filed a patent application for technology that rather than have the the user viewing artificial light emanating from a screen, the device will project light directly onto the wearer's retina [3]. The holy grail with all of these technologies is to create an environment where the user interacts with virtual and physical objects in a natural, seamless way. It appears the goal of many of these technologies is to make the virtual objects as similar to the physical objects in the immediate environment, to the point that the user is unable to distinguish between the virtual and the real in his or her interactions.

Certainly the cameras that are incorporated into AVFDs are an obvious point of concern regarding the technology. Denning, Dehlawi, and Kohno [4] conducted a small experiment of reactions bystanders have to cameras and recording devices. Their work revealed that the newness and unfamiliarity of these devices caused bystanders to view them differently from other recording devices such as mobile phones. Their mock recordings took place in a cafe and many bystanders thought the researchers ought to be required to get permission before recording. Some bystanders showed an interest in a (hypothetical) device that would block such recording.

While Denning et al. focused solely on recording components, AVFDs certainly will contain other familiar components such as GPS. Some of the ethical concerns we raise are not new to AVFDs; however, the nature of these concerns change when these technologies are combined into a single device with proposed components of AVFDs such as holographic projectors and retina projectors. Often times promoters of these technologies speak of the advantages the individual user of the device will experience. There seems to be little analysis of both the potential disadvantages to the individual user and almost no analysis to the impact these devices might have in larger groups and on social structures. We address some of these ethical concerns here.

Friedman and Kahn [5] examined augmented reality using seven human values they predicted would be important for understanding the ethical import of AR. In our earlier paper, we explored three of those seven: psychological well-being, physical well-being, and privacy. In this paper, we will concentrate on the remaining four values in Friedman and Kahn's list: deception, trust, informed consent, and ownership. We will also draw from the additional values that Friedman and Kahn suggested in a subsequent paper [6]. They include freedom from bias, universal usability, autonomy, identity, calmness, courtesy and environmental sustainability. AVFDs (especially future devices) will embed several of these values.

Our discussion focuses on four groups of stakeholders involved with AVFDs: developers, users (both individually and collectively), non-users who are in sight of users, and society as a whole. We will use the term "developers" in a broad sense, meant

to include at least designers, software engineers, and managers of the companies making these devices. We will call non-users who can be physically (not virtually) seen by AVFD users at "the watched."

2. DECEPTION

Deception is like the tango – it takes at least two, a deceiver and a deceived. In [7] we considered deception to be "an intentional, successful attempt by developers to deceive users, and a misapprehension by people other than the developers." Consistent with Lynch [8], deception requires a misleading act that is "*willful or non-accidental*. So, X deceives Y with regard to f only if X willfully causes Y to fail to believe what is true with regard to f." It is important to note that deception is not inherently bad. As we noted in [7], developers regularly hide implementation details from users to make the user experience more familiar (e.g. the use of a the folder and file metaphor for the file system). We called this a benign deception. In this section we describe three possible deception relationships that we think are both likely and ethically significant with AVFDs. All three relationships involve users, but one of them includes developers, and one of them includes the watched. AVFDs seem to strain Lynch's requirement for deception that the act be willful or non-accidental in that a user of an AVFD may make a willful act of using one, but have his/her reality become so intertwined with a virtual component that the possibility of willfully deceiving someone is no longer a conscious choice.

A more interesting take on deception for this application might be that of Mark Wrathall [9]. Wrathall offers insight into deception as a perceptual experience. Wrathall writes, "In the genuine perceptual experience, the phenomenal character of things corresponds to the way things actually are. One then accounts for deceptions by treating them as the presentation of a certain phenomenal character in the absence of the objects necessary to make that presentation true" [9]. He goes on to explain that "when we are deceived, it's because the thing really looks like what we take it as." So deceptions, in this sense, have to do with misperceptions. It is how we view the world and how the world is presented to us [7]. This raises an interesting question for the case of AVFDs where a genuine perceptual experience includes not only phenomenal character of things but also the virtual. Everyone's perception of the same object may be different because of what is virtually added. We would not call this a misperception but rather an augmented one. So, how can we tell if an augmented perception is a deception? In a certain sense, AR is all about fooling the user's eyes and brain. So where do we cross the line? Great care should be taken to help users be discerning consumers of this new information.

2.1 Developers May Deceive Users

AVFD developers have several kinds of power over users. First, the developers know many technical details about the devices, details that are not obvious to most users. Because of this information (and power) imbalance, developers could deceive users about the capabilities and sophistication of the AVFD and its algorithms. This kind of deception would not be distinctive to AVFDs, but is common to all high tech devices. However, the nature of AVFDs, the intimacy of changing what people see, might increase the ethical significance of this particular technology deception.

The ancient slogan "seeing is believing" [10] illustrates another way that AVFD developers might deceive users. Should developers succeed in engineering the AR experience in such a way that augmented reality is indistinguishable (or nearly so) from physical reality, users might be deceived into believing in the physical existence of what they see, even though it is not physically present. In the case of devices that display light directly on the user's retina, the intention to deceive cannot be eliminated from the nature of the AVFD. The user cannot distinguish the two different sources of light. It will take other cues for the user to determine the virtual from the physical.

Regardless of whether a virtual object is a holographic image or being displayed directly onto the user's retina, the developer takes on additional responsibility for the veracity of any information attached to the object. Either purposefully, or carelessly, developers could deliver bogus information to users. It may very well be that users who see that information called up instantly and effortlessly into their visual space will be inclined to give that information the benefit of any doubts about the information's accuracy. One way to mitigate this concern would be to make it obvious to the user the nature of control that she has over information. Yet one of the developmental difficulties is determining a convenient way for a user to provide input into an AVFD. Shortcomings in this feature lead to more control for the developer and less for the user. Therefore great care should be taken to help users be discerning consumers of the information they are perceiving in order to mitigate the potential for developers to routinely deceive the users.

2.2 Users May Deceive the Watched

In considering how AVFD developers may deceive users, we concentrated on AR outputs to the user's eyes. In considering how AVFD users may deceive the watched, we also consider AR visual inputs, real time video taken from the user's viewpoint. The potential for privacy invasion was one of the reasons Google Glass users were not universally welcomed into public spaces [4, 11]. Users, recording members of the watched, might explicitly or implicitly lie about their actions or intentions.

In addition, users might misrepresent what they are seeing via their AVFD. We can envision many scenarios in which a user either has or might have information that non-users do not have. A user might be asked about that information, or a user might volunteer it. Either way, the user might misrepresent the presence or absence of the requested information, or misrepresent the information. "Yes, I can see that…" could be used as a method of establishing authority and seeking the power of information (whether the information is true or false). Rather than create an atmosphere of trust, these potentials for deception create one of distrust and uneasiness.

2.3 Users May Deceive (Other) Users

One of the interesting aspects of AVFD systems is the potential for multiple users (who will probably have to be using similar, if not identical, systems) to interact in the virtual space overlaid on their individual physical views. So, for example, we are told that we will be able to play virtual chess, or laser tag, with each other. But it does not take a great deal of imagination to anticipate that some AVFD users who share virtual space with other users could rig the common virtual experience to their individual advantage. For example, one laser tag participant may find a way to have the game unfairly slanted to his or her advantage. There are numerous ways a virtual poker game could be used to cheat opponents. Users might purposefully share inaccurate information (for

example, information about other people in the room) that would be displayed on nearby AVFDs.

A single user of an AVFD may choose to deceive him/herself. Someone may choose to adorn him/herself with opulent jewels or even keep a long dead pet close at hand. As virtual worlds and the physical world become increasingly blended, questions about what is real will begin to change. Each of us can create our own blend of physical and virtual to create our own realities. Underlying assumptions about all of us sharing the same reality will no longer hold.

3. INFORMED CONSENT AND AUTONOMY

Medical informed consent [12] implies knowledge of the intended intervention, awareness of possible risks and benefits, and an explicit declaration of agreement that the procedure go forward. Applying this idea to the use of AVFDs, several aspects already discussed seem relevant. Consistent with the sensibilities of the subjects of Denning et al.'s experiment [5], there is a case that both AVFD inputs and outputs should be considered for informed consent. First, a user who uses an AVFD to record images or audio should do so only with the consent of people included in the recording, particularly if the recording is going to be shared. This is further complicated when the recording includes holographic images that appear to be a real part of the physical space. A simple, uniform method for describing the nature of the recording, what is allowed and what is not allowed and the actual obtaining of consent from the watched, especially in large crowds, seems to be no simple feat.

Furthermore, if a developer or a user is responsible for changing another user's virtual space, it should be clear to the affected user that this change is taking place. Surely some such changes would be well known by the users involved; if a user bought a virtual chess program, and the developer delivered an appropriate set of virtual objects for the players, no formal informed consent would be necessary as it is implicit in the product. But if a developer or a user X controlled the virtual space in such a way that all watched individuals were scanned, and otherwise private information appeared in X's virtual view, then the watched individuals should be asked for their consent, or it should not be done. We can imagine scenarios (for example, in an emergency room) where watched patients might be willing to give such consent to medical staff with AVFDs. But we can also imagine some patients in an ER who would refuse consent. Either way, it should be an option, not a requirement, for treatment.

We can envision situations in which people would waive AVFD informed consent For example, some AVFD enthusiasts might want to gather and experience each other's virtual manipulations. If fellow users were trusted, or if enthusiasts did not care about the consequences of giving up their control of virtual space, they might mutually agree to a common license (among themselves) for a wide-open experience. As long as such agreements are explicit and mutually agreed upon, we do not see an ethical problem.

We can also envision scenarios in which someone was coerced into using an AVFD. The coercion could be economic, where as part of your job requirements you had to agree to training with an AVFD. The coercion could be legal; for example, probation could be granted only if a prisoner was willing to undergo AVFD therapy, therapy that was designed to induce revulsion at certain triggering situations. In these types of cases, authorities (commercial or governmental) may reason that the greater good (of a corporation or a polity) trumps the need for voluntary informed consent. We are suspicious of such reasoning, and we contend that great care should be taken when forcing AVFD experiences on to individuals.

While it is likely that in general people will not be forced to use AR devices, we can envision certain contexts where such use may be encouraged or even required, for example at work or in school. An AR device that can "pin" holographic objects in the real world and can allow users to interact with both virtual and physical objects simultaneously seems to offer a potentially valuable learning environment. A student in a class which requires interaction with a pinned hologram would seem to have little choice but to acquire and don an AVFD. Using such a device as a classroom tool is not necessarily ethically problematic if all students have access. However, "having access" may be more complicated than simply having a device to use; some students may not be able to benefit from an AR device. Blind students are an obvious example, but some sighted students might have adverse reactions to an AR device, including headaches or dizziness; how will such students be treated if an AR experience is a required part of a curriculum? Teachers have a tradition of guiding students' learning in similar ways. However, issues of autonomy do creep into this situation. We need mechanisms to determine the level of control each student should have. The teacher and the school will also exhibit some level of control over the experience, with one or the other potentially having complete control over each student's use of the device. As a collaborative and learning tool, it may be useful for students to see the interactions and the results of interactions that other students initiate.

4. OWNERSHIP

Several ownership issues arise surrounding AVFDs. First, will AVFDs be owned (like most computer hardware) or leased (like much proprietary software)? We assume that the AVFD hardware will be owned, but that much of the software will be leased. Proprietary software is likely not to be readily accessible for users or for the watched; therefore, there may be interest in having at least some AVFD software be free or open source software (FOSS). We will not reprise the arguments for and against proprietary and FOSS solutions here, but this is a venue where those arguments will again play out, affecting the balance of power between developers and users, and to a lesser extent between users and the watched.

In previous sections we pointed out the possibilities for deception and informed consent situations having to do with AVFD users recording images and sounds from the watched and from other users. This aspect of AVFDs can be viewed as an ownership issue: who owns my recorded image and voice? Legally, particular instances of this argument may turn on where the AVFD is deployed. If the recording takes place in a public space, then the watched may not have a presumption of privacy; if the recording takes place in a space that is not legally designated as public, then there may be a presumption of privacy. However, we suspect that an ethical analysis would be more restrictive of a user's "right" to the use of the watched's images and voices. For a more complete discussion of AVFDs and privacy, see [1].

The issues of ownership of devices and recorded images for AVFDs are interesting, but closely related to issues with previous devices. Graham, Zook, and Boulton [13] demonstrate the power that comes with one augmented reality technology, Google Maps,

by demonstrating how Google shows and describes places differently depending on the language one uses to view a particular place. A more distinctive ownership issue for AVFDs is: who has legitimate claims to the virtual space (what the users see)? We assume that a user should have some claim to that space, since it is his/her device, and since his/her eyes and visual cortex are most immediately impacted by the virtual image. However, the developers of the device work to design and deliver that virtual environment, and they might also make a claim of ownership; the developers clearly do have control, especially initially, on that virtual space. If some AVFD applications require real time Internet sharing (similar to what gaming systems use for multiplayer games), again that virtual space is claimed by both developers and users.

This sort of sharing also suggests a need for open standards. Proactive work on how virtual objects and experiences are to be represented and shared will allow for users with different brands of devices to be unencumbered by those differences. There is a need in the AVFD arena for the same sort of frictionless interaction that we experience while texting, making phone calls and sharing photos.

In cases where both developers and users may have possibly legitimate claims to ownership, we think it is vital for the participants to have explicit agreements about the ownership of the virtual space. It may be that in particular applications (such as shared AVFD games), users will be content to relinquish control in order to enter into a group experience. In other applications (for example, a surgeon using AVFD during an operation), users may demand a much higher degree of control, especially when they are responsible for critical decisions based partly on information delivered by an AVFD. In both these cases, the stakeholders can act ethically, but only when the agreements are explicit, appropriately detailed, and understood by all parties.

One virtual space of particular interest is that surrounding existing physical objects. The Artvertiser project started by Julian Oliver [14] seeks to "improve reality" by placing virtual art over advertising in public spaces through the use of AVFDs. While the virtual art is visible only to the wearer of the AVFD, it does "prevent" the wearer from seeing the advertisement on the billboard. An advertiser might argue to the AVFD developer that such an ability ought to be blocked on the AVFD. Since so much software on portable devices is largely supported by advertising, this sort of feature might lead to a decrease in economic support of software available for AVFDs or an increase in the price of that software. On the other hand, there is no clear argument that one ought to be subjected to advertising in public spaces. Even without AVFDs, people can avert their eyes. Yet, the intriguing question remains, should someone be allowed to own the visual experience in a public place?

Closely related to that question is perhaps the most important aspect of AVFD ownership--that of an individual's ownership of his/her own perception. In some sense, donning an AVFD allows someone (or something) to radically alter what the individual perceives. This temporary surrender of control has analogs in other technology. When we see a film at a theatre, when we watch television, and when we listen to an iPod, we are giving control over one or more of our senses to a machine and the sociotechnical system of which that machine is a part. But the distinctive mixture of physical and virtual that is delivered by AVFDs may be seen as a qualitatively greater surrender. And if it becomes commonplace to make that surrender on a daily, or even continuous basis, then part of who we are, and much of what we see, will be "owned" outside of ourselves. That is a major ethical issue with power at its core..

5. TRUST

AVFDs are artifacts that mediate our perception of reality. According to our Object Oriented model of Trust [15], they would fall under the category of human to human trust mediated by electronic means. There we state: "The people who design, develop, or deploy a computing artifact are morally responsible for that artifact, and for the foreseeable effects of that artifact. This responsibility is shared with other people who design, develop, deploy or knowingly use the artifact as part of a sociotechnical system." [15] What is the impact on trust?

There are two trust relationships that must be considered: trust between users and developers; and trust between users and other individuals (some of whom may be users themselves, and other individuals who are not users). Both the developers and users must take on moral responsibility for the artifact. That is, developers of AVFDs should have as an accepted goal: examination of the effects of that artifact on society and performance of their functions with the appropriate standard of care. A subgoal here would be transparency: developers being honest with others about the capabilities of the device. Users who trust developers will buy their products and use them with confidence. However, if the user performs certain actions based on the trust he/she has in the artifact, and if that trust is misplaced (i.e., the developer is manipulating the end-user and does not have the user's best interests at heart), then there is a violation of trust [8]. In the second trust relationship, individuals must trust that users in public are employing the device in an ethically acceptable way.

Another issue of trust involves epistemic trust. How do we know what we know from our perceptions through AVFDs? Can we trust what we perceive to be true? Judith Simon says that "trust and knowledge are fundamentally entangled in our epistemic practices. Yet despite this fundamental entanglement, we do not trust blindly. Instead we make use of knowledge to rationally place or withdraw trust. We use knowledge about the sources of epistemic content as well as general background knowledge to assess epistemic claims. Hence, although we may have a default to trust, we remain and should remain epistemically vigilant; we look out and need to look out for signs of insincerity and dishonesty in our attempts to know" [16]. This statement could apply to the user's relationship with the developer. It is more difficult to trust what we see as true when the virtual and real are entangled and our world is mediated through a device. How does what we know impact what we perceive and conversely how does what we perceive impact what we know? The answer to these questions will affect whether we trust what we see through the AVFD.

6. IDENTITY

In addressing issues of identity, we note that AR devices may help individuals establish their own identities. There is the potential for a deep blending of the physical and virtual self. In the physical world, people use jewelry, body piercings, tattoos, and ear lobe gauging to distinguish themselves and establish at least part of their identity. People use posts on Pinterest, FaceBook, Twitter and Instagram to create a virtual part of who they are as individuals. AR devices open up the possibility of pinning these sorts of identity-creating virtual items to one's physical self, so that anyone with a compatible AVFD will see the pinned objects

when you are viewed. It could become similar to having "virtual jewelry," with vendors competing to offer increasingly sophisticated decoration that shares all of the properties of information such as being easily and quickly updated. Those viewing someone who is virtually decorated through an AVFD will see that person as part physical, part holographic and perhaps be unable to distinguish between the two.

Of course, that is the ideal. A person ought to have autonomy over her/his identity, yet the AVFD through which the person is being viewed may be owned by someone else. At the very least there is the opportunity for the owner of the AVFD to use a different virtual accoutrements on a person than those selected by the watched. The possibility of decorating *others,* especially without their consent, seems fraught with difficulties and potential abuse. The potential problem becomes even more pronounced in a group setting. This seems to be especially true in a setting, such as a school, where bullying takes place. This technology opens up new avenues for cyber-bullying.

Ethical principles surrounding identity seem to collide with ethical principles surrounding public spaces. In the case of a public space there is a reasonable argument that an individual can choose to use an AVFD to obscure or replace an advertisement in a public space. In some sense the person lays claim to the visual space between the AVFD and up to, but not including the advertisement. On the other hand, when the AVFD user is viewing another person in a public space, the user's right to control their own visual experience comes up against the watched's identity and autonomy. As in so many questions about technology and people, power relationships are clearly important. Inasmuch as AVFDs empower individuals to thrive, there is a positive effect; inasmuch as AVFDs are used to enhance the power of those already powerful to the detriment of the less powerful, there is a negative effect.

7. ETHICAL CHOICES

In order to elaborate some of the ethical choices to be made with AVFDs, consider the following scenario: developers have set up a system that includes multiple users and the developers themselves to interact using interlinked AVFDs, sharing a physical and a virtual space together. Two examples of such a situation could be a developer, a surgeon, and a group of medical students inside an operating room; or a group of gamers and a developer in an outdoor setting playing a first person shooter game. What can we say about the actions of the developers in this situation?

First, the developer has at least two kinds of control in these situations: first, the developer controls the initial configuration of the system, including what the users will see (virtually), and how much control each user and the developers have over those virtual images. (In this paragraph, we will use "images," but in many AVFDs, there could also be sounds added.) The second kind of control is real time, after the AVFDs are deployed, and the users are inhabiting the same physical and virtual space. In a move toward simplicity, the developer might decide that no one's virtual images take precedence and thus block everyone's. This option certainly detracts from the value of AVFDs. At the other extreme, the developer could allow everyone's virtual images to be seen. This also seems to detract from the value of AVFDs as such an experience would be visually cluttered and noisy.

For a more nuanced look, consider the interests of a developer D and a user U (who is not a developer). Assume that D wants to associate certain virtual images V to U so that anyone looking at U with one of the AVFDs in the system will see V (virtually) as well as U (physically). Several different situations arise:

1. U does not like some aspects of V, and objects to D, either before V is shown to others, or after V is shown to others. Whose preferences are likely to take priority? That probably depends on the situation, and on the power relationships outside the AVFD system. For example, in the medical situation, the surgeon will probably have a great deal to say about how s/he is presented to others, but a medical student might not have any say. In a gaming situation, users might have some latitude for some aspects (for example, they might make up a gaming ID that is virtually attached to them), but not for other aspects (for example, the game may insist on projecting their current life force). Deciding whose preferences *should* take priority, the ethical question, will be situation specific; however, we contend that developers should negotiate these kinds of issues early and often during development and deployment.

2. Now assume that two users, U1 and U2, have the power (granted by the developer) to change virtual images associated both with themselves and with each other. Again, conflicts can occur when one of the users "decorates" the other with images that the decorated user finds objectionable. We think it is central to the ethics of this situation what agreements the users entered into when they joined in the AVFD system. If they agreed to subject themselves to this decoration by others, then they probably may not have much to complain about. If U2 objects to U1's decoration of U2, U2 can try to negotiate with U1 to remove or change the decoration, or U2 can withdraw from the system.

3. In both case 1 and case 2, the issues can be framed as informed consent. Thus a system that informs U2 of U1's desires and allows U1 to either consent or not seems to be called for. This option has the positive of forcing an interaction. It does not seem to impinge excessively on U1. In the end, it opens the opportunity for collaboration, allowing both U1 and U2 to potentially thrive. This approach impinges on the developers, forcing them to design an entire system for this sort of exchange to take place. This is an interesting case for the ownership issue as well. Certainly, one ought to expect bullies and trolls to avoid this sort of thing, and it would be unreasonable to expect this sort of system to not be hacked. If the software were FOSS, then it would be easy for the bullies and trolls to avoid informed consent. Proprietary software, on the other hand, would make that more difficult.

Traditionally, the question of ownership of the space between an object and someone's eyes has not been called into question. AVFDs have the potential to force us to consider that question. It opens up new opportunities for individual freedom for AVFD users (e.g., one can avoid being bombarded by advertisements in public spaces), and also potential hazards for the watched who could be seen not as they physically and virtually project themselves, but rather as the AVFD user would like. This is a collaborative environment of public and private, virtual and real. Developers, and the systems that they produce as part of AVFDs, will have an important role to play in the environment that surrounds these devices.

8. CONCLUSIONS

The phrase "I can't believe my eyes" is meant to say that something is extraordinary, surprising, and unexpected. But if it becomes commonplace not to believe our eyes due to AR devices and policies that allow others to control what we see, we think that we will be engaging in a risky socio-technical experiment. We contend that such issues should be debated now, not after AR devices become commonplace.

9. REFERENCES

[1] Grodzinsky, F. S., Miller, K. W., and Wolf, M. J. 2014. Augmented reality in your eye: Google Glass, Space Glasses, and beyond. *In CEPE 2014*.

[2] Microsoft. 2015. Microsoft HoloLens. http://www.microsoft.com/microsoft-hololens/en-us.

[3] Abovitz, R., Schowengerdt, B. T., and Watson, M. D. 2015. Planar waveguide apparatus with diffraction element(s) and system employing same. http://www.faqs.org/patents/app/20150016777.

[4] Denning, T., Dehlawi, Z., and Kohno, T. 2014. In situ with bystanders of augmented reality glasses: perspectives on recording and privacy-mediating technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2377-2386. DOI=10.1145/2556288.2557352 http://doi.acm.org/10.1145/2556288.2557352

[5] Friedman, B., and Kahn, P. H., Jr. 2000. New directions: A value-sensitive design approach to augmented reality. In *Conference Proceedings of DARE 2000: Design of Augmented Reality Environments*. ACM, New York, NY, USA, 163-164.

[6] Friedman, B., and Kahn, P. H., Jr. 2003. Human values, ethics, and design. In J. A. Jacko and A. Sears (Eds.), *The human-computer interaction handbook*. Mahwah, Lawrence Erlbaum Associates, NJ, USA, 1177-1201.

[7] Grodzinsky, F. S., Miller, K. W., and Wolf, M. J. 2015. Developing automated deceptions and the impact on trust. *Philosophy and Technology*. 28, 1, (Mar. 2015), 91-105. DOI=10.1007/s13347-014-0158-7.

[8] Lynch, M. P. 2009. Deception and the nature of truth. In M. Clancy (Ed.), *The Philosophy of Deception*. New York: Oxford University Press, 188-200.

[9] Wrathall, Mark A. 2009. On the existential positivity of our ability to be deceived. In M. Clancy (Ed.), *The Philosophy of Deception*. New York: Oxford University Press, 67-81.

[10] Ammer, C. 1997. Seeing is believing. *The American Heritage Dictionary of Idioms* Houghton Mifflin Harcourt Publishing Company.

[11] Kelly, H. 2013. Google Glass users fight privacy fears. http://www.cnn.com/2013/12/10/tech/mobile/negative-google-glass-reactions/index.html.

[12] Dictionary.com 2015. "informed consent," in *Dictionary.com Unabridged*. Random House, Inc. http://dictionary.reference.com/browse/informed consent.

[13] Graham, M., Zook, M., and Boulton, A. 2013. Augmented reality in urban places: contested content and the duplicity of code. *Transactions of the Institute of British Geographers*. 38, 3, 464-479.

[14] The Artvertiser. http://theartvertiser.com/.

[15] Grodzinsky, F. S., Miller, K., and Wolf, M. J. 2012.Moral responsibility for computing artefacts: 'the rules' and issues of trust. *Computers and Society*. 42, 2, (Dec. 2012), 15-25. DOI=10.1145/2422509.2422511

[16] Simon, J. (2010), The entanglement of trust and knowledge on the web, *Ethics and Information Technology*. 12, 343-355.

First Dose is Always Freemium

Kai K. Kimppa
Postdoctoral Researcher
Information System Science
University of Turku
kai.kimppa@utu.fi

Olli I. Heimo
Project Researcher
Technology Research Center
University of Turku
olli.heimo@utu.fi

J. Tuomas Harviainen
Postdoctoral Researcher
School of Information Sciences
University of Tampere
tuomas.harviainen@uta.fi

ABSTRACT
In this paper we look at three different groups of games. The traditional payment methods for games, although they do have their problems, are typically less problematic from ethical perspective than their more modern counterparts. Payment methods such as lure-to-pay use psychological tricks to lock the player to the game. Whereas pay to pass boring parts or pay to win just use game-external mechanics to make the play easier, and thus intent to, and have consequences other than at least many of the players would want to. This paper is a first stab at the topic from a Moorean just-consequentialist perspective, and in future papers we intend to compare a wider range of philosophical methods, payment methods as well as look into empirical data on players views on the topic.

Categories and Subject Descriptors
J.4 [**Social and behavioral sciences**]: *Economics, Psychology*

K.4.1 [**Public Policy Issues**]: *Ethics*

K.4.4 [**Electronic Commerce**]: *Payment schemes*

K.7.m [**The Computing Profession / Miscellaneous**]: *Ethics*

K.8.0 [**Personal computing / General**]: *Games*

General Terms
Design, Economics, Human Factors, Theory, Legal Aspects.

Keywords
Games, Ethics, Economics, Payment models

1. INTRODUCTION TO COMPUTER GAMES' PAYMENT METHODS

"Hate the game – not the player"
– Anonymous

Two late examples from the game industry illuminate the difference we want to highlight in this paper: Rovio Entertainment with its Angry Birds franchise (comprising not just of the games[1] themselves, but from anything from pet toys through soft drinks to amusement parks) and SuperCell with games such as Clash of Clans. The introduction of mobile market places – especially, Google Play, AppStore, and Windows Phone Store – has made this possible. The business models of these two companies, however, differ greatly. Rovio Entertainment makes games that you either download for free and suffer the advertisements or pay for and get the game, and that is it. Add-ons, or rather sequels and even fully autonomous different style games such as Angry Birds Space come out every now and then, but follow the same model for sales. A large part of the revenue is made from the sales of fan items and other accessories, many of which just show game icons or themes (e.g., a logo on a shirt), but do usually not feed back into the game experience beyond possibly enforcing a loose neotribal identification as players of that game (see [1]). As the accessories are outside the scope of this article, we will not be taking a stance on those. On the other hand, SuperCell's Clash of Clans as well as their other games like Boom Beach are free-to-play; but to succeed, the player really needs to buy additional in-game enhancements, effectively making the game a pay-to-win game; as is wittily pointed out by those who consider themselves actual fans of computer gaming.

The growth of the game industry has been phenomenal. Game consoles, as well as 'traditional' computer games have started to sell more and more – digital distribution services such as Steam and similar at least partly explain this, but also the fact that the age group who grew up with games is now adults – and have money – and are introducing their own offspring to games explains some of it. Clearest examples of these are Wargaming.net's World of Tanks and World of Warplanes, as well as games from World of Warcraft style initial payment and monthly fee, to freemium (i.e. "free-premium", game is free but by paying the player receives some additional features or advantages) converted BioWare's Star Wars: the Old Republic. In these high-budget MMOs the "first dose" is clearly free, but to advance in the game one is strongly pushed towards paying a monthly fee to actually play (and enjoy) the game. Wargaming.net clearly promotes other in-game purchases e.g. better equipment and ammunition in their games. BioWare restricts most of the end-game content from the freemium players.

[1] Unless specifically noted, "game(s)" in this article is used to mean *digital* games rather than any wider range of games – this just to increase readability of the paper.

Also, 'casual games' that come with our social media applications, such as Farm Ville with Facebook and others, explain a part [2]. And, as hinted in the first paragraph, the explosion of the mobile marketplace in the form of tablets and multimedia mobile phones. All these have made the gaming industry a considerable part of our current economy. Unfortunately, with growing economies, the bad is often along with the good. Many scholars of games, from Huizinga [3] and Caillois [4] onward, have stated that one key trait of games is that they stand apart from everyday life and what happens in the play has no consequences outside it. The darker sides of real money trading show that this is not always true.

Players see the intrusion of monetization directly to their gameplay as problematic. For example, freemium players interviewed by Paavilainen, Alha and Korhonen [5] indicated as their key problems "boring gameplay" and "interrupting pop-ups" (followed by four interface issues and then one on the necessity of recruiting one's friends to play). Both of these are elements the skipping or removal of which is a well-known way for getting game revenue. This indicates that monetization systems may be a direct disruption to enjoyable play, and thus problematic. To clarify which payment methods are good and which are not, we will introduce Moor's [6] seminal paper on "Just consequentialism and computing", and look for a solution from his thought that both the intention and the consequences of the solutions we select for our applications must be ethically sound; else we are not designing an ethical piece of software.

Moreover, the time and effort spent in a freemium game is a value to the users themselves. In many such games, one is able to obtain the same things with either real-world money or with time and effort spent on playing. Generally, players who have money tend to value their time, while poorer players are willing to spend more of it in order to avoid paying with money [7]. Yet still the freemium game can easily be modified, upgraded or even the whole logic of the game or its payment methods can be changed. According to Leavitt [8] and Nurminen and Forsman [9] that can and will affect the whole information system, the game, the developers and the players. How, depends on the change, and usually is not known. The alternative ways of acquiring things, the real-money exclusivity of some things, and the changes to them both take place may well be (and probably are) revenue logics for the game publishers [7]. Players, however, do not necessarily agree on them being appropriate (e.g. [10]). Some love the new options; others see them as an unfair intrusion into the gamescpace. This issue though, according to Heimo, Kimppa and Nurminen [11], makes it also an ethical issue.

2. JUST CONSEQUENTIALISM

As Moor [6:65] points out, since computers – and especially their software – are malleable, they often develop at a far faster rate than our legal or social traditions do (see also [12]). Thus, we are left with policy vacuums: we do not know what is right and what is wrong in a given situation. This is especially true in the current computer game markets: new methods to profit from games pop up faster than anyone – be it researchers, game developers, law makers or the public – can keep up with.

To help these groups assess the morality of games, we shall, in this paper investigate a selection of different payment methods for games – whether pre-purchase, purchase, in-game sales, unofficial services, add-ons or other – through James Moor's [6] just consequentialist framework. We are of course aware – as was Moor [6:65] – that theoretically consequentialist and deontologist theories do not mix. None the less, in every-day ethical decision making most of us use a mix of one sort or another of these (among possible other theories, such as rights-based or virtue ethics). Should an application fail either requirement – the intention not being honourable, or the result not being beneficial to the target groups – the application is morally suspect, according to Moor [6].

A typical example of an immoral application would be spam. It is immediately obvious that neither the intent nor the consequences of spam are moral. The intent is to get people to buy things they neither need nor want (e.g. breast or penis enlarging pills or other "remedies"), and the consequences are even worse, as the prescribed remedies practically never deliver. A similar situation can be found in the aforementioned free-to-download Rovio game Angry Birds. The banner ads in the game are hardly something the player wants (so they fail the intention criteria), nor do they deliver what they promise (after all, advertising is just legalized lying; "show me an ad, and I will show you a lie" still holds as true as ever); and, it appears that in later incarnations Angry Birds has even resorted to pay-to-win tactics for revenue[2].

From a social perspective, real-money trading may appear to be an unfair advantage, or even amount to cheating [5, 13, 14]. In many freemium games that is nevertheless an intended part of the design and a key part of their publishers' revenue models [13, 15]. A central challenge for assessing the ethical aspects of real-money payments towards especially winning in a game like World of Tanks is therefore that since the designer-enabled RMT is embedded in the functioning logics of the game itself, it can be considered a part of the expected infosphere of the game. From the direction of information ethics, if the game supports a certain kind of behaviour, it has to be considered appropriate *within the constraints of that game*, even if our values outside of that play would not find the actions (e.g., paying to win) ethical [16]. Once we accept the social contract to play, we accept that within the temporary reality of that play, the rules and ethics may not be the same. Or do we?

3. IS IT RIGHT OR IS IT WRONG TO PAY OR NOT TO PAY?

3.1 Traditional payment models

For the purposes of this paper, videogames can be grouped in three different categories (and several subcategories) according to the method of payment (see also, [7, 17]). The first group is the "traditional" group which falls into categories "Pay once", "Pay periodically", "Freeware" and "First dose".

The first model we look at is the traditional off-the-shelf payment model, because it was for a very long time the most successful – and with shareware and freeware the 'only profitable commercial' – model in gaming distribution and thus we compare every other model to this.

We call the model **"Pay once"**. In this model the customer expects that when they pay at the counter for the game, they get the whole game, and there are no other, hidden payments afterwards. The customer purchases the entire product, plus

[2] We would like to thank the anonymous reviewer who pointed this out – as well as reminded us to check all of our examples for accuracy. Also, for this particular example, see "Mighty Eagle" option. http://angrybirds.wikia.com/wiki/Mighty_Eagle

possibly the right for some free updates to it. This also typically is the case (although there are exceptions, to which we will get later). The main aim of the game producer is clear: produce a game, get paid for the game, hope the customer is happy with the product so that they will like the game and hopefully praise the game to their peers so that others will buy the game as well and that the customer is happy enough with the product to buy other games from the producer later. Of course there might be other motives for the producer as well, but these motives are typically benign ones, such as that the customer will enjoy the game and have great time playing it. Thus, the *intent* of the producer is to make money and have a satisfied customer; both quite laudable goals. Thus, the first rule of a benign intent is satisfied.

The customer expects to get a game that they can start playing directly after they get the game home and install it on their computer. Whether the game then later is in need of patches, which are expected to be provided by the game producer at no additional cost, or whether add-ons or later versions of the game are sold are typically irrelevant to the original purchasing decision, as it is the game itself that the customer is buying. Now, whether the customer is actually satisfied with the content of the game is outside of the scope of assessing consequences in this particular case, as that depends on the expectations on the content, rather than the expectations of the payment model. Thus, the consequences *from the payment model* are good in the sense that the customer gets exactly what they expect: the full game and nothing less or more, or more likely, what they expect to get – a problem common with any marketing.

The off-the-shelf game does not of course need to be an *actual* off-the-shelf product. If a game online is sold with the same motif, whether directly to the customer (so that a customer copy remains) or through a portal (as long as the licencing agreement is actually equivalent to buying an off-the-shelf game), this makes no difference to the analysis. The situation does, however, get more complicated when purchasing from services such as Steam or Origin online stores, as the end-user licences often stipulate additional conditions, such as if the service is no longer available, the customer may lose access to the games or one-sided conditions on 'misuse' (e.g. modification without 'permission') or withdrawal of service without notification. In these two later cases it is clear that the intent of the, in this service provider, not necessarily game producer is not as benign as it is when they just sell the product and then it is the customer's choice what they do with it. If such conditions are set the good *intention* does not satisfy, and thus the seller is not following just-consequentialist model.

Thus, the traditional off-the-shelf game satisfies both requirements of good intent of the actor, the game producer, as well as good consequences for the targets of the action, the customer, i.e. the player.

Freeware is of course quite clear, in the pure form. The purpose of freeware distributors is typically not to fool the player, as they are giving that particular game or the version of the game away for free without any strings attached, nor are the consequences from the player's perspective problematic – they get the game and get to play it. The game of course could be an advertisement for the next version of the game (or another game by the same producer), and as advertisements are *always* lies, and as lying is wrong, this may pose a problem intentionwise. None-the-less, the game itself is not intended to lure the player into playing for that game. Next versions may however require payment, and thus some freeware games can be considered to be – if not for the intention, at least for the consequences – lure-to-pay games.

Traditional **shareware** games (or, these days demos of games that can be downloaded to try the game) follow the model of the off-the-shelf game with the minor exception that first parts of the game are made available for free. This is also clear from the outset; the customer is not fooled into thinking that they would be getting the whole game for free, but are aware that only the taster is free, the rest has to be purchased if they want to play it [7]. Again, clearly both the intention and the consequences are benign, if the off-the-shelf standard model is otherwise followed. Surely the problem with marketing stands still as in "pay once" model.

Many current games, especially mobile games fall into the **lure-to-pay** category. Almost all freemium games are in this group. In these games the idea is to – in various ways – offer the gamer a fairly large amount of the game for 'free', just spend your time playing it. And then spring the psychological trap: sunk-cost fallacy; I have invested *so* much time on this game, it must be meaningful to me, thus, to advance, I am willing to pay in one form or another to get the next part out of it (see e.g [18]). Either some parts of the game are excluded if the player does not pay for the rest of the content or advancement is cut short if they do not. The games – like World of Tanks – can even be changed to be lure-to-pay games after the fact, changing the rules (of course within the EULA, but who reads them) after players have invested hundreds of hours in a game they believed to be free-to-play.

The intention of lure-to-play is to trap the customer. Either from the beginning or changing the rules on the fly (i.e. changing the game type to lure-to-pay). The consequence manifests itself through the sunk-cost fallacy; the player cannot evaluate the actual value of the game independent of it. Thus they are willing to pay even in a situation where they actually do not value the game but just the time spent with the game. Therefore both the intention and the consequences are at least problematic if not upright malicious.

"Pay periodically" has been used traditionally with B2B deals but due the growth of the Internet it expanded to the video game business. In this model players typically pay for a month, three months, half a year or a year at a time is by default not problematic from a just-consequentialist perspective as typically an online server service is included, which both causes the service provider costs and offers the player an online environment in which to play the game with others for added value compared to "pay once" model.

The intention from the service provider is clear: they are providing a service for the player which they think provide the player additional value. From a consequence perspective the player gets exactly what they pay for – the online environment, the added value from being able to play with their online co-players and, what is most important, they know what they pay for, as even though online environments are offered also for 'free' (never *actually* for free, though), the player is typically not fooled into believing that they get something else than what they expect. Yet there are some problems shared with the "lure to play" model, such as the time, money and effort invested to the game that can cause a psychological dependency for the games which is clearly more problematic – as an intention as a consequence – than any other model described before.

3.2 Pay While Playing

The second group is called "Pay while playing". This group contains only the subgroups "Pay to win" and "Pay to pass

boring". In-game money (or in-game golds) is a method of mimicking reality in the games and to give the player the resources to allocate between different options. Allocating the money optimally is in most of the key for victory and a big part of the game itself and in most games the requirement to advance in the game. Thus the game company selling the in-game money can ease up the gaming and either a) move player through "boring parts" of the game where the players should "farm" – repeatedly do same tasks to gain resources – in order to advance or b) shift the balance between the players in such a way that the one who pays gets advantage over others. Companies also use "offline progress" as a tool: many games recharge energy needed for activity while a player waits, or e.g., have growth times for plans, and players can skip such downtimes with micropayments [5].

As the former – **"pay to pass boring"** – is only a shortcut to more meaningful playing experience. It can be provided either by the game provider or by an external party, such as Chinese gold miner (see e.g. [19]). A key thing that makes games games is that they contain artificial limitations on activities, so that the activities themselves become more enjoyable (e.g., in boxing, one does not try to get the opponent down by any means necessary, but rather rises to the challenge of doing so within the constraints of set rounds, limits on what may be done, and even wears gloves that weaken the blows; [20, 21]. Play to pass boring, however, makes them even more inefficient, in a less exciting manner, and seeks to get people to pay so that they can get to the enjoyable parts. Already the fact that in many environments this service is provided illicitly by a third party raises some reservations on its morality. The aim from the game company perspective is to offer a 'service' which makes the game easier for the player – through resources outside the game. This is of course true for a third party service, such as World of Warcraft levelling service provided by Chinese or Mexican miners as well. The purpose on both accounts is not to provide the player additional value but rather rip their money so that they can pass boring parts of the game (either intentionally inserted, if the game provider, or areas which *seem* boring but actually make the player better at understanding their character and thus would provide additional value if they went through the trouble of learning their character, by a miner).

The latter, **"pay to win"**, gains straight advantage for those who are willing to pay. It is a model clearly immoral from a just-consequentialist perspective. Whoever has enough money wins, or at least those who have money, have a chance to win. No gamesmanship necessary; just *bribe* the system to win – except when someone else is bribing too, when the gamesmanship (or more bribes) becomes a necessity. But moreover, if you do *not* bribe, you lose. Always (or at least often enough to be the norm). The aim of the game provider is to create a situation in which the player has invested enough time and effort to feel the pressure to finalise the win by using money (see also: "lure to pay"), and consequently, they cannot win if they do not invest money on top of time and effort. Many players tend to hate this approach [22] and, interestingly, even many game design professionals (mostly those who do not work with freemium model games, though) find it unethical, if taken to the extreme where it is impossible to win without paying [13]. It is a financially risky strategy, as it alienates players not willing to engage in constant RMT. It can sometimes be a triumph as well – for example, a player of World of Tanks able to beat by skill alone others who are trying to pay and win may gain significant pleasure from that fact.

The problem with these is common where the money of the customer is clearly linked with the imaginary money in-game. Thus the money – and effectively spending it – is connected to the gaming experience of the player and all the other players of the game. Hence the intention is the same: the experience must be made decent enough for the player to get caught and to utilise the sunk-cost fallacy to promote the urge for the player to see the game and its experience in such scale that the investment to skip boring parts or to increase the odds of winning are worth the money invested to the game. The consequence is even harsher; the game must be made so that those who are willing to pass the boring parts are numerous enough that is, the game must be made boring enough thus lessening the entertainment value of the game altogether.

Even so, the advantages of the payment in pay-to-win model must be made so vast that the player willing to pay must get a reasonable advantage thus limiting the odds of success for all the other players. Thereby the game must be made more unfair and yet more boring for these models to work properly and only by paying more can one get a reasonable experience. Finally the main problem is that when one is paying in games like this it is not like paying once or paying a monthly fee. Instead, one is paying an unpredictable amount and therefore one cannot foresee the amount of money one needs to pay to pass all the boring parts or to succeed in the game. The latter one of course has yet another problem: it can be forced to an arms race.

3.3 Content and Access

Third group is the "Content and Access" group with the sort of obvious subgroups "Content" and "Access". These methods are more likely to be quite contemporary, experimental – or even futuristic. The most common methods of gaining these are e.g. *new gaming content, access to use some options in the game, add-ons, downloadable content (DLC), possibility for multiplayer and removal of unwanted content such as advertisement*, all through payment but perhaps not so obviously – limitedly [7, 22]. Thus the content and access are more or less the two sides of a same coin where the one more or less follows the other, e.g. access to new content.

Downloadable content is a complex issue from a just consequentialist perspective. If the DLC is actually created after the sold content and meant to offer more, but not critical content to the game, it passes muster both from the perspective of intent and consequences – the intent is to give the player an option to buy more material to extend their gaming experience (in many ways similar to user generated content (UGC)) and it does enhance their gaming experience. However, much DLC is created during game creation and the game is handicapped by removing it before release; making the game, if not unplayable, at least clearly diminishing the playing experience, and thus even the intent, let alone the consequences is immoral.

New gaming content sold within the game is typically a fairly clear issue however – it is just ripping the player off. The player is unaware of the option, and is – suddenly – awakened to the option of buying ingame content that they were not aware that they needed to purchase to be able to complete the game.

On the other hand, offering a user generated content option for the players is *the* optimal just consequentialist gaming addition – giving the players tools to extend the game is intentionwise the thing to do, and the consequences could not be better from the players perspective! It also appears to be an increasingly viable business strategy for publishers, despite the fact that monetizing content created by third parties can be rather tricky [7]. That Steam had to remove its "paid mods" option very quickly by no means spells the end of that market.

Much of what players trade between each other consist of scarce goods – if a virtual item is sold, the seller no longer has it. In contrast, much of the contents that can be purchased via real-money trading from the publishers count as information goods, and their value in use does not diminish even when others make similar purchases [22]. It can, however, nevertheless become lower, due to two factors: firstly, ubiquity lessens novelty – if everyone has access to lots of gold ammunition in World of Tanks, or everyone gains access to a certain location in an MMORPG, it becomes the norm rather than the exception. Secondly, purchases and possessions are often about prestige. They are status symbols, in both games and outside of them [23]. If one is able to buy the same item with money, which other players spent hours or days of works on, the value of the item decreases for both.

A particularly curious issue is the removal of advertisements. From an ethical perspective, they are like intentionally boring grinding content, but more intrusive and furthermore external: an unnecessary distraction to play, an interruption from outside of the game as artefact proper, intended to inspire people to remove the problem (with money) so that they can concentrate on playing [7].

Some of these games are marketed as "Free-to-play", whereas it is more complicated. There are multiple free-to-play models which differ quite a bit from one-another. These range from free, as in there are no fees, through buying cosmetic enhancements such as prettier skins to buying game-enhancing items/skills to making progress quite slow if the gamer is not willing to pay to excluding access to certain areas or levels if the player does not pay.

It is of course obvious that many of the cases are either borderline cases or combinations of the two or even more. Thus it seems that the modern yet undefined field of monetary transactions within electronic games are – because of both the contemporary and yet unmolded field of gaming – harder to analyse with the moorean theory at hand, or at least to compare with models that have been unmodified for decades. Therefore more research and time for these kind of payments is required.

4. SOME GAMES ARE FAIRER THAN OTHERS

When we look at the following methods to pay for a game or its content through Moor's [6] just consequentialist framework, it becomes soon apparent that some methods do not satisfy at least one of the requirements – and that is enough, according to Moor, to point out that there is something suspicious in the method.

The first two models are clearly unproblematic from a just-consequentialist position. They give the full game content for free, with no strings attached; if the game producer then sells eye-candy as extra, this in no way typically affects the game engine itself.

None-the-less, the other 'traditional' free-to-play models, such as game is fully available, but making progress slow if no payments are made are more complex. In the case of slowed progress for those who do not pay, this needs to be clearly understood by the player from the beginning, and even if it is, it still often changes the game balance between players who pay and those who do not. The 'bored grinder' learns their character better than the 'shortcutting payer', and this creates imbalances in playing skills. If the game requires cooperation, it can be seen to be unfair towards the players who put in the hours if they cannot tell who are taking short cuts, and thus do not necessarily pull their weight in joint play.

There is a clear difference between paying to win, paying to shortcut and paying for extra content, in both monetization and in ethics. A "whale" (much-spending player) buying a "harpoon" (expensive content or item aimed for that specific player segment; [7]) may present its own ethical questions. Yet as noted by Sicart [16], those are from a systems perspective non-problematic, being a part of the game, even as such spending may carry heavy real-world consequences and dilemmas [13]. It is at the borders of the other monetization types, particularly competitive games with a pay to win option built into them, where we really see the murky waters.

5. CONCLUSIONS AND FURTHER RESEARCH

Based on the previous it is clear that some ways of getting a payment for games are more justified than others and it is obvious that the different payment methods generate different kinds of results in accordance to the method used as well as the overall architecture of the marketing and branding. Thus the customer – or in some advertisement-based games the 'product' – can be seriously taken advantage of by the game developer.

The more modern the payment method, the more it seems that there are problems looking at it through the just consequentialist framework; the harder it is to say whether the intention or the consequences are beneficial or harmful – and the harder it is, the more suspicious it seems. One could even say that there was a kind of righteous naivety in *'the olden days'* of game development: no psychologists were used to find more and more compelling ways to 'rip' money from the players, yet others might argue that the economic window of video gaming was not fully developed. Whether either of these is true remains to be seen.

However, the previous is a look through a theoretical framework – we still do not know what actual gamers think about the various payment methods. Yet again a different method should provide us different results and thus the "ethical truth" in its theoretical form is hidden deeper within the analysis itself. Therefore more analysis on various points of view must be acquired. Moreover the theory requires some empirical data. Thus in a future paper, we (together with other colleagues) will look into this issue and compare the results from empirical data to the framework used in this and possibly other papers.

REFERENCES

[1] Tyni H. & Sotamaa O. (2014). Material Culture and Angry Birds. In Proceedings of DiGRA Nordic 2014.

[2] Juul, J. (2010). A Casual revolution: Reinventing Video Games and their Players. Cambridge, MA: The MIT Press.

[3] Huizinga, J. (1955). Homo Ludens: A Study of the Play Element in Culture. Boston: Beacon.

[4] Caillois, R. (1961/2001). Man, Play and Games. Urbana: University of Illinois Press.

[5] Paavilainen, J., Alha, K. & Korhonen, H. (2015). Domain-Specific Playability Problems in Social Network Games. International Journal of Arts & Technology special issue on Advances on Computer Entertainment.

[6] Moor, James H. (1999), Just consequentialism and computing, Ethics and Information Technology 1, 65—69.

[7] Lehdonvirta, V. & Castronova, E. (2014). Virtual Economies: Design and Analysis. Cambridge, MA: The MIT Press.

[8] Leavitt, Harold J. (1964), Applied Organization Change in Industry: Structural, Technical and Human Approaches. In Cooper, W.W.,Leavitt, H.J. and Shelly, M.W. (eds.): New perspectives in organizational Research, p. 55—71. Wiley. 1964

[9] Nurminen, M. I. & Forsman U. (1994), Reversed Quality Life Cycle Model. In: Hu-man Factors in Organizational Design and Management - IV, 393—398, Elsevier Science B.V., North-Holland, Amsterdam, 1994.

[10] Paavilainen, J., Hamari, J., Stenros, J. & Kinnunen, J. (2013). Social network games: Players' perspectives. Simulation & Gaming, 44(6), 794-820.

[11] Heimo Olli I., Kimppa Kai K. & Nurminen Markku I. (2014) Ethics and the Inseparability Postulate: How to make better Critical Governmental Information Systems, Ethicomp 2014, Paris, France, 25-27 June, 2014.

[12] Lastowka, F. Gregory & Hunter, Dan. (2004). The Laws of the Virtual Worlds. California Law Review, 92(1), 1-73.

[13] Alha, K., Koskinen, E., Paavilainen, J., Hamari, J. & Kinnunen, J. (2014). Free-to-Play Games: Professionals' Perspectives. In Proceedings of DiGRA Nordic 2014.

[14] Lin, H. and Sun, C.T. (2011), "Cash trade in free-to-play online games", Games and Culture, 6(3), 270-287.

[15] Hamari, J. and Lehdonvirta, V. (2010), "Game design as marketing: how game mechanics create demand for virtual goods", International Journal of Business Science & Applied Management, 5(1), 14-29.

[16] Sicart, M. (2009). The Ethics of Computer Games. Cambridge, MA: The MIT Press.

[17] Hamari, J., & Järvinen, A. (2011). Building Customer Relationship through Game Mechanics in Social Games. In M. Cruz-Cunha, V. Carvalho & P. Tavares (Eds.) Business, Technological and Social Dimensions of Computer Games: Multidisciplinary Developments (pp. 348-365). Hershey, PA: IGI Global.

[18] Hamari, J. (2011). Perspectives from behavioral economics to analyzing game design patterns: loss aversion in social games. In Proceedings of CHI'2011 (Social games workshop), Vancouver, Canada, May 7-12, 2011.

[19] Kimppa, Kai K. and Bissett, A. K. (2008), Gold Farming, in Ethicomp 2008, University of Pavia, Mantua, Italy September 24-26, 2008.

[20] Suits, B. (1978). The grasshopper: Games, life and utopia. Toronto: University of Toronto Press.

[21] Salen, K., & Zimmerman, E. (2004). Rules of play: Game design fundamentals. Cambridge, MA: MIT Press

[22] Harviainen, J. T. & Hamari, J. (2015) Seek, Share, or Witthold: Information Trading in MMORPGs. Journal of Documentation 71(6).

[23] Lehdonvirta, V., Wilska, T.-A. and Johnson, M. (2009). Virtual consumerism: case Habbo Hotel. Information, Communication & Society, 12(7), 1059-1079.

Wilma ruined my life: how an educational system became the criminal record for the adolescents

Olli I. Heimo
Project Researcher
Technology Research Center
University of Turku
olli.heimo@utu.fi

Minna M. Rantanen
Graduate Student
University of Turku
minna.m.rantanen@utu.fi

Kai K. Kimppa
Postdoctoral Researcher
Information System Science
University of Turku
kai.kimppa@utu.fi

ABSTRACT
The information system StarSoft Wilma used to track and report on the adolescents' behaviour at school can cause problematic situations. These problems manifest themselves in various ways: many of the markings in the system are either wholly unnecessary or at least questionable in nature. This is made the adolescents (and some guardians/teachers) resent the system. In this paper these side-effects are looked through and compared with an analysis of posts in the Facebook-group 'Wilma Ruined My Life'. As conclusions we claim that the system can create an atmosphere of fear and suspicion amongst the students: resembling an Orwellian or panopticon-like environment which might undermine the students' ability to become full and capable members of an open democratic society.

Categories and Subject Descriptors
J.1 **[Administrative Data Processing]**: Education

J.4 [**Social and Behavioral Sciences**]: Psychology

K.4.1 [**Public Policy Issues**]: Ethics

K.6.5 [**Security and Protection**]: Physical security

K.6.4 [**System Management**]: Centralization/decentralization

K.7.m [**The Computing Profession / Miscellaneous**]: Ethics

General Terms
Your general terms must be any of the following 16 designated terms: Management, Design, Security, Human Factors, Legal Aspects.

Keywords
Student Information system, Ethics, Pedagogics, Psychology, Adolescents

1. INTRODUCTION
"You may leave school, but it never leaves you."
- Andy Partridge

The Finnish primary and secondary school information system StarSoft Wilma is the interface-part of StarSoft's school information system group which is developed to ease information processing and communication within schools and between schools, students and their guardians. With Wilma the different interest groups, e.g. students, teachers and parents or other guardians (henceforth just guardians) can communicate, share information and view timetables.[1] It is usually seen as "a school journal", a notebook traditionally used for communication between school and home, but it extends to a be-all-end-all system for storing information about schoolwork[2].

Wilma was implemented during the 2000's to large amount of Finnish schools and is used in most of the Finnish municipalities as one of the primary tools in teaching. Many of the students however view the system often as a "student criminal record" [2, 3, 4, 5, 6]. There is even a Facebook group "Wilma pilasi elämäni"[1] (eng. Wilma Ruined My Life) where students discuss Wilma and post their teachers' comments. The group has over 60 000 members (over 1% of Finnish population)[5]. It is worth noting that each year in junior high (grades 7, 8 and 9) which is the group discussed in this article is approximately the same size as the membership of the group. Even though it is clear that not all members of the group are from junior high – after all there are members such as the authors of this article – the majority of the members are either current or former targets of Wilma.

2. WILMA AND YOUTH
Imagine that your boss uses Wilma and reports your hourly behaviour there. Imagine that you have worked a lot in the morning and then you start to feel a little bit tired after lunch and you do not work that efficiently for couple of hours. Then imagine that you get home and your mother calls you and tells you that you are lazy and you have to work harder. How would you feel? How would you feel if that happened more often? This little example is reality for many adolescents. You as an adult might not be as disturbed by this indirect feed-back that you got, but someone more prone to not receiving critique well could take it a lot more personally.

[1] https://www.facebook.com/WilmaPilasiElamani

In this article we concentrate on students that are in their adolescence – the age between childhood and young adulthood, attending grades 7, 8, and 9 – ages 12 to 16. Adolescence is often thought to be a life phase filled with turmoil – teenagers are going through set of physical, psychological and social changes in which they have to adapt to. They are in the middle of developing for example their cognitive skills, identities and morals.

In this chapter the effects of Wilma in adolescent development are considered. It is worth noticing that the intention of this chapter is to raise some issues that should be considered when Wilma is evaluated – not to argue that Wilma has some specific effects on the development of adolescence. Studying the latter in full would need the study of its own. Thus, this chapter is merely a glimpse of theories of development and how Wilma can be seen as a part of that process.

In the field of psychology there are many different theories about development of adolescents. For example Erik H. Erikson's [7] psycho-social development theory describes that adolescence is all about establishing an identity. According to this theory, establishing an identity as a task of a life phase is completed when one finds an identity that feels genuine and makes one feel complete. If identity forming does not succeed the person becomes conflicted by the many roles and does not feel the identity obtained to be genuinely theirs.

However the idea which Erikson presented about the identity forming process is interesting when thinking about influence of Wilma and the way adolescent use it. According to Erikson [7] forming an identity is a social and dialectic process between individual's needs and environmental interactions. The identity development process is mainly social since it is based on how the social environment defines a certain individual and how the individual interprets that. Thus, the way that the adolescent is being defined for example by the guardians, teachers and the peers affects a lot on how successful they are in their identity development.

Evaluation is a big part of schools, but Wilma has made evaluation of students more constant and easier. Although Wilma has been intended to be used as a conduit for constructive feedback (see e.g. [8]), many entries are merely critiques or notions about behaviour as presented later in this paper. Although Wilma entries are not meant to be personal critique towards the students and their developing identities, they can interpret them as such. The students can see the entries as something that reflects the teacher's idea about them and adapt that as part of their identity although entries are only one way to give feedback about their work in school. This is especially harmful if the feedback is only negative and does not give instructions how to act to get better in school.

Wilma ruined my life Facebook group is an example of how negative feedback is turned into positive. Instead of being publicly ashamed about these notions, adolescents are seeking positive attention from their peers. In the group negative feedbacks are often seen as humoristic. In 2013 the existence of this group even lead to a competition of funniest Wilma entries in primary schools, which lead to banning schoolkids from Wilma in the Helsinki area [9]. However, this action has not ceased the flood of Wilma entries to Facebook and has made Wilma a communication tool between teachers and guardians by excluding the schoolkids.

Wilma is not all about negative feedback – it also gives teachers an opportunity praise good behaviour. Alas negative feedback is over presented in Wilma systems. For example in a Facebook group "Tieto- ja viestintätekniikka opetuksessa/ICT in Education"[2] one teacher was overwhelmed by the fact that their Wilma contained 8 negative and only 2 positive options for feedback. Positive options included only very active behaviour in school, so he felt that there should be more options so that bigger part of students could gain positive feedback. This teacher understood the importance of both negative and positive feedback and saw constructive feedback as a way to give his students a system that does not only punish them but makes them feel appreciated for their efforts to work better.

Nurmi [10] argues that development during youth is a process in which the adolescents steer their life, get feedback about decisions made and develop an idea about themselves through that feedback. The majority of adolescents are on "a positive track" – they set goals, find ways to achieve them and gain a feeling that they have achieved something. When they fail, they are able to adjust their goals and think new ways to achieve them. Unfortunately this is not the case with everybody – some tend to fail and instead of adjusting their goals and the ways to achieve them they start to use defensive mechanisms such as blaming something or someone else for their failure. This can lead to a vicious circle where the adolescent does not take responsibility and continues on the same path. This can lead to poor success in development and to behaviour problems.[10]

If the feedback is mainly given through Wilma – in which some have no access to – and it is unconstructive, how the adolescents keep on the "positive track"? When schoolkids are excluded from Wilma giving the constructive feedback about the behaviour in school becomes a duty of the guardians. The guardians have to interpret Wilma entries and try to create a narrative around them which they understand even though they do not know the whole story. By excluding adolescents from this information exchange about themselves, the schools are actually making adolescents more depended on their guardians instead of supporting their process of becoming more independent.

Although intentions behind Wilma might be good there are some issues with its use. Reporting everything about an adolescent's day in the school does not seem an effective way to either support their success in school or the development of the adolescent. There is also a possibility that the adolescents grow into being subjects of an Orwellian society where their actions are recorded and used against them.

3. WILMA AS AN INFORMATION SYSTEM

The nature of the information is private but yet the information is delivered to the guardians of the adolescent. Thus the private acts of the adolescent are not private but something shared with the adolescent and their guardians. Moreover the markings in the Wilma system are but a mere glimpse of the whole: it is something the teacher sees proper to report. Even though there obviously are guidelines on how to use the system and what should and should not be reported, in the end the decision lies with the reporter: the teacher. Hence it is relative both to the student and the teacher alike what actions from the day are reported – or is anything reported at all. Thereby equal treatment of subjects – the adolescent – is nearly impossible.

Unfortunately, the system supports situations in which when the teacher can misuse the system to punish a student for an act which they have not committed, for example over either a disagreement

[2] https://www.facebook.com/groups/237930856866/

with the guardians or because a quality/trait/feature of a student that the teacher for one reason or another disagrees with. This feature is not done only by writing a report to the system but moreover by marking and using those markings as a proof and stigmatisations and thus a qualification for yet a harsher punishment. On the other hand, positive comments of other teachers can lead to situations in which teachers are more likely to make positive comments or leave out negative comments because there is a previous record of positive behaviour reported by their colleagues.

One main difference with the electronic information system (compared to traditional analog one) is the way the information is stored. When the information is stored in digital format it can only be copied with photocopier or by digitalizing it. With larger centralized information systems it can be copied by few clicks of a mouse – by a municipality worker or a cracker alike.

Yet the information security has been left to the municipalities in which the level of information security can vary according to the skills of the IT personnel. E.g. StarSoft tells in their FAQ that the municipalities/schools can use their own security keys which are not graded as proper by the browsers. Moreover the storing of Wilma records – including their upkeep and eventual destroying – is left for the problem for the municipalities.[1] Thus the records can be stored indefinitely and the information can be stolen (i.e. copied) and used against the then-adolescent in their later life.

Moreover the system is used as a 'criminal record'. There still lies the problem that this criminal record is not administered by the central government but the IT-supports of the city governance. Thereby some of the information can be stored and later accessed by parties that should not have access to it. Youngest members of the parliament of Finland are only 24 years old and thus – if this holds true in next elections – Finland will have MPs who actually have had their actions recorded to Wilma. In Finland – as well as globally – some of the governmental information has leaked to the press and this kind of information could feed the yellow press heavily for a week or two – just enough for the tabloids to make their profit out of it. Thereby – and for all the other reasons – it is important to protect the privacy of the adolescents' – now and in future.

4. HOW WILMA RUINS LIVES

There are guardians who refuse to use the system for various reasons: they feel the system is too much of a watch dog, think their kids should be given more leeway in their life, and that only serious issues should be reported, Moreover, according to them, when it should be reported, it should be done personally, not through an impersonal system. Many of the received notifications are also rather irrelevant; notes such as a student being a bit late for class or having missed class even when being sick and guardians are – or should be – aware of this. In any case, most of these issues would solve themselves by default, without any need for marking them in any system. Others are also worried about privacy issues. The default in some school districts for Wilma use is not to separate between guardians when showing them things. Normally, this is not a problem, but for example in the case of separated guardians, items that should not be available to both guardians (e.g. the other guardian's personal replies to the teacher) should be separated in the system as well.

Since school is a major part of the social environment of adolescents, it should support positive development. Wilma however seems to involve the guardians in their children's everyday life, although the adolescents are usually trying to separate themselves from their guardians. Feedback given through the system could support the development, but based on observations and inquiries the feedback is often not constructive (see next chapter). Feedback from peers is also important and adolescent can seek it from the social networks. Negative feedback from teachers can become something that is valued amongst peers and it can encourage adolescent to misbehave. One example about this kind of behaviour is "Wilma ruined my life" group in Facebook.

From the Facebook group the following things can be derived:

1) Wilma is a big issue amongst the youth
2) Many markings are quite meaningless
3) Teenagers intentionally misuse the system

Many of the markings the students have posted to the group are quite meaningless: "you did not quite concentrate" or similar minor misbehaviour. Yet, more interesting is the way how children brag about their more uncommon markings such as "You went to hide your moped from the police immediately after the class had started"[3] or "Yelled to everyone she hid the notebook to her ass, when I asked to write words to the notebook". Thus, it seems, the system to control the kids has turned out to be a proof for "funny" behaviour – at least for some.

The misuse of the system is a concern regularly raised in public discussion (see e.g. [2, 3, 4]) and has sprouted a large amount of discussion in Finnish discussion boards. This is not only due to the controversial use of the system, but the clever teenagers who have found a way to rebel against the system through social media. The Facebook group seems not only to be a forum to discuss and criticize the information system, but also to be a form of self-actualisation, a place to brag for the teenagers and compare who is the most creative and most daring.

Jeremy Bentham's brother, Samuel introduced the concept of the Panopticon in late 18th century. [11] The idea was to create a prison in which the inmates could be unobtrusively followed all the time, without the inmate knowing when, if at all, the guards were looking at what they were doing. They could do it all the time, not at all, or, what is more important, whenever they wanted. Foucault [12] modified the concept to include the whole society. We are all being potentially watched all the time, and this, when we are aware of it modifies the way we behave in the society; we do not do the things we otherwise would – and would want to. Both of these are problems Wilma introduces to the students.

The most central problem is the excessive control of the pupils. Wilma causes both Benthamian [11:172-173] Panopticon for the students in that they can never know what is written of them in Wilma or, especially who looks at what is written about them, as well as a Foucaultian [12] Panopticon, which makes them change their habits in fear of being observed and reported. Even though the students also flaunt their misbehaviour, they do it as a rebellion towards a system many feel to be too intrusive and unfair.

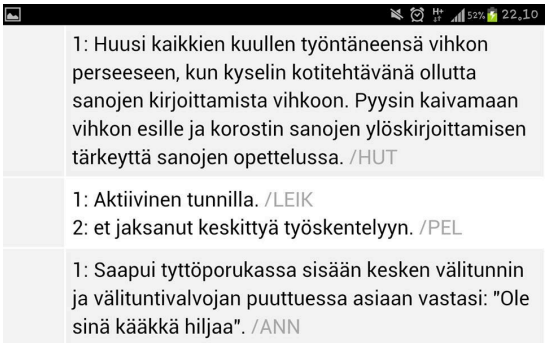
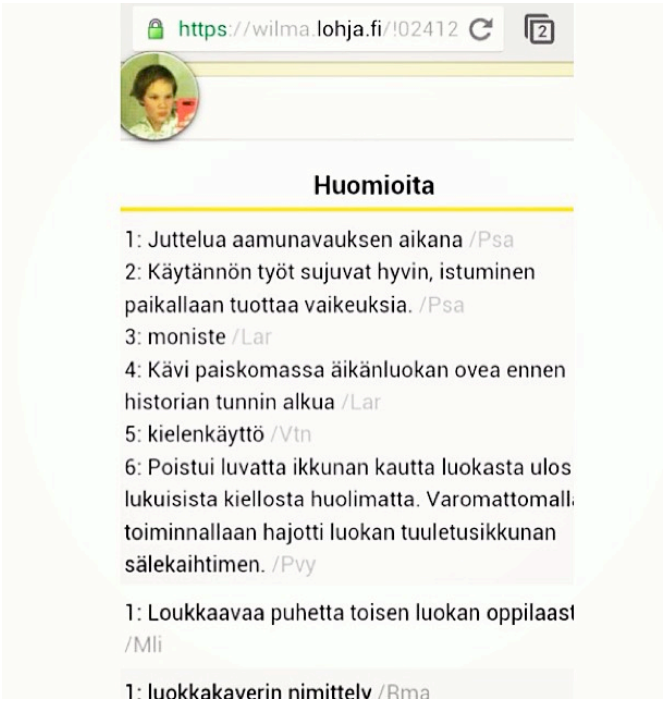
Thus the Panopticon for children is a serious issue for the next generation of adults. If one as a child learns to submit to being observed and reported even for most ridiculous reasons with unfair treatment, it is much easier to not question that kind of treatment as an adult.

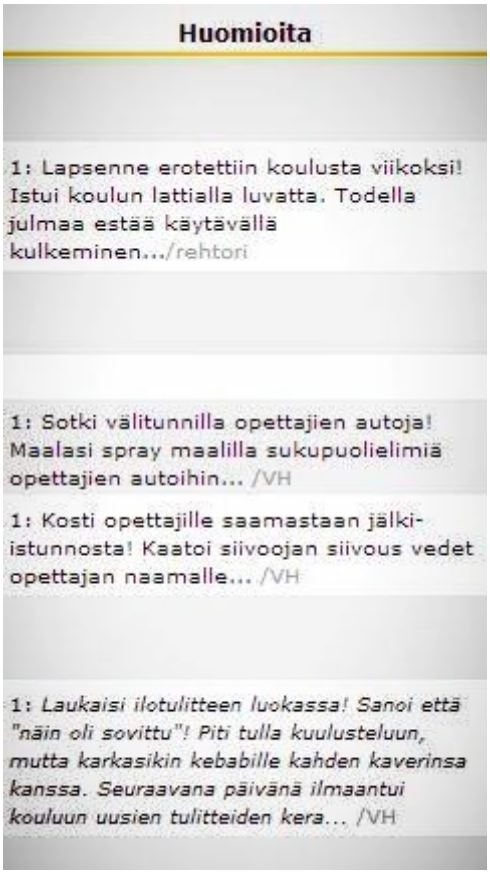
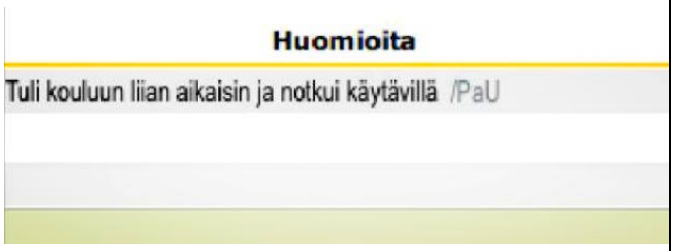
[3] Tuning mopeds to illegal motorcycles is a regular hobby amongst teenagers in Finland.

5. RESISTANCE TO WILMA

As mentioned before, the Wilma system has spawned some resistance in social media. The following table (Table 1) contains some demonstrative screenshots.

Table 1: screenshots and translations from Facebook.

Translation	Original
1. Yelled in front of everyone that she put the notebook into her arse, when I requested to write things down. I requested to take the notebook and emphasized the importance of writing words up in the learning. 1. Active during class 2. You did not have the energy to concentrate during the class 1. Came with other girls inside during the break and when teacher intervened responded: “Shut up you old hag.”	
1. Talking during opening of the day. 2. Practical work goes fine, sitting still is difficult. 3. Print-outs 4. Went to bang Finnish-class door before the history class. 5. language 6. Went out from the classroom window without a permission against many warnings. With dangerous activity broke the classroom window’s blinds. 1. Disrespectful talk about other class’s student. 1. Calling classmate with names.	

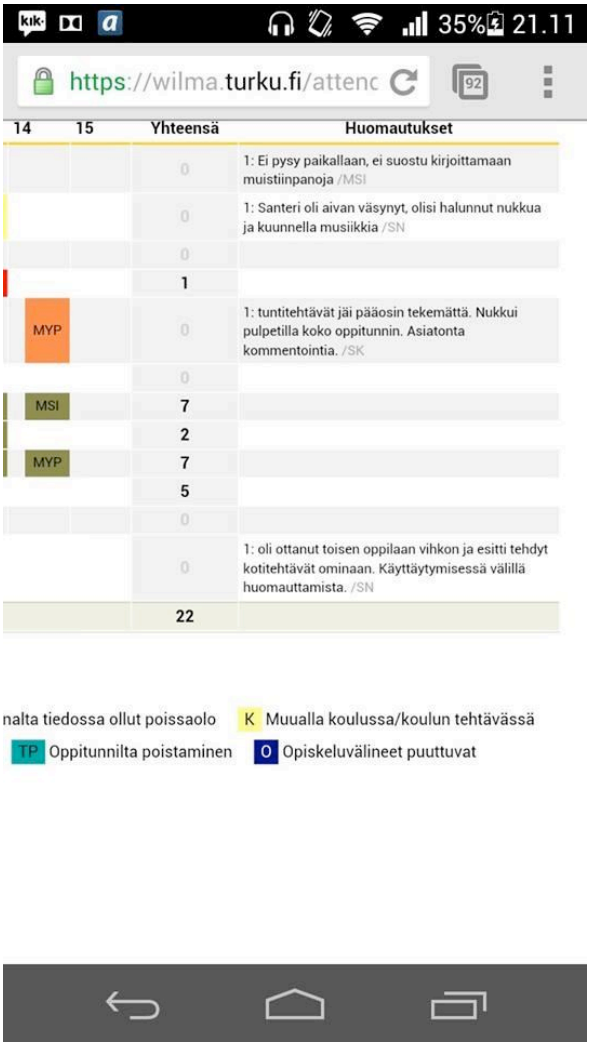
1: Your child has been expelled from the school for a week! Sat in the school floor without permission. Very cruel [sic!] to block the movement in hallways… 1: Messed teachers' cars during the break! Painted with spray paint [grammar error] genitals to teachers' cars. 1: Avenged the detention to teachers! Poured cleaner's cleaning waters [another grammar error] to teachers face… 1: *Fired a firework in classroom! Said that "this was agreed upon"! Should have arrived to questioning but ran off to have a kebab with two friends. The next day arrived to the school with new fireworks…*	
Came to school too early and hanged out in the hallways.	
1: Also the notebook is missing. 1: Where are the books? 1: Did not do anything during gymnastics. 1: books not with [grammar and spelling…] 2: Equipment missing. 1: math mid-test: 8+ [in a scale from 4 to 10] 1: Did nothing during gymnastics. Slept in the stage.	

1: Can't stay still, doesn't agree to write notes.

1: Santeri is too tired, he would have preferred to sleep and listen to music.

1: assignment [sic] mostly undone. Slept at the desk the whole class. Inappropriate commenting.

1: had [sic] taken another student's notebook and presented the answers in it as his own. At times some inappropriate behaviour.



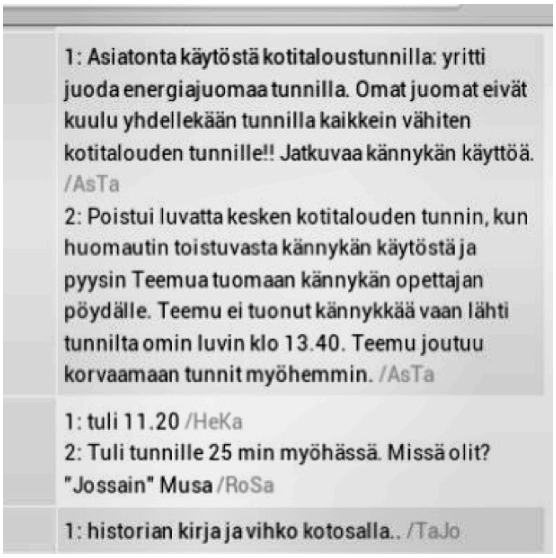
1: Inappropriate behaviour during home economics class: tried to drink an energy drink during the class. Own drinks are not appropriate during any class least of all in home economics!! Continuous cell phone usage.

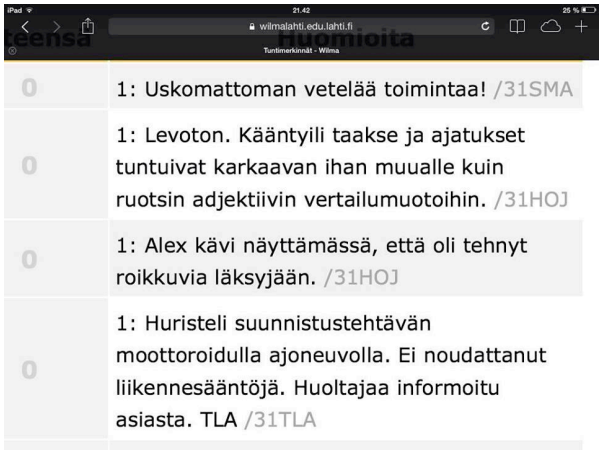
2: Left without a permission during home economics class, when questioned about repeated cell usage and I asked Teemu to bring the cell to teacher's desk. Teemu did not bring the cell but left the class with his own permission only at 13.40. Teemu will be required to repeat the classes later.

1: Arrived 11.20

2: Arrived 25 minutes late. Where were you? "Somewhere" Music [sic].

1: history [sic] book and notebook at home.. [sic.]



1: Unbelievably lazy action! 1: Restless. Turned around and thoughts seem to run somewhere else than Swedish adjective comparisons. 1: Alex showed some hanging homework assignments. 1: Drove to the orienteering with a motorized vehicle. Did not obey any traffic rules. Guardian has been informed about the issue.	
I got grounded for a month because of this =D 1: According to eyewitnesses showed her breasts to the classroom boys. In interview denied the event, according to the boys was paid a coke for doing it. "And it was not a coke but a fanta :D :)"	

Many – if not most – of the notes are somewhat unimportant; in others, rebellion is showing – the students are just playing with the system and using it to strengthen their status within their social group. It must be remembered that these are mainly copied from Wilma ruined my life Facebook group, and thus they do not represent full image of Wilma notifications. It is worth noting that the relevant notifications in Table 1 are not only the obvious ones, such as student calling the teacher "an old hag", but also those which seem pointless, such as "Where are the books?"

These messages give a kind of taste on what kind of information is delivered to the guardians of "bad boys and girls". As mentioned before, teachers seem to keep their notes really short and unconstructive although StarSoft has clear instructions about communication between teachers and guardians. Ben Furman [8] has developed a list of instructions to teachers how to communicate using Wilma. Furman for example states that when negative feedback is given teachers should tell how they wish students to behave in the future and what the benefits of that behaviour are. In examples given teachers neglect this advice in whole by only stating what has happened. Furman also recommends that teachers should create confidence in future success and respect the expert position of guardians when it comes to their children.

Examples given in Table 1 are also examples how teachers misuse Wilma. They use it as "criminal record" instead of using it to communicate with guardians. By following simple instructions of Furman [8] Wilma could have more positive effects on development of adolescents – they would know what is expected from them and that there is hope to get better. Also guardians would know how they can help their children to behave in the manner that teachers require. If majority of Wilma notions are only statements the information drawn for them is only a list of what adolescents should not do in school.

Listing banned actions does not support the development of adolescents. It only teaches them to obey rules that can be irrational. For example students are not allowed to be late from school but they are not allowed to be early either.

Also the resistance should not be a surprise. As Volkman (2014) states, *"[w]hen some central political authority dictates culture, opting out of the culture expresses a rejection of the moral authority of political power. When subcultures, tribes, and individuals rely on the moral authority of their own narratives, any authority left to a wider politics will be as neutral umpire rather than arbiter of the true, the beautiful, and the good. […]The efforts of states, communities, corporations, schools of thought, and various other suprapersonal entities to micromanage and author the lives of individuals are inevitably met with*

resistance from individuals and the churning and seething intercourse of their various overlapping subcultures."

Thus the control has in this case – and yet again – met the resistance. However it seems that the rebellion seems to be only online raging which – after the number of headlines in yellow media – seems to have gone back to the underground. Therefore only the rebellious ones – those who collect reputation online and those who support them – can really be said to be active against the system. The updates in the group (few per month) are not a big amount of rebellious activity and the supporters (60 000) seem to be quite passive as well. It should be obvious though that the private groups of adolescents are more active in the matter. Therefore this rebellion seems to have failed its' true purpose – if any – while the majority of the subjects just settle to mumble in their own private groups.

6. CONCLUSIONS

As shown before, Wilma is a tad bit controversial due the problems reported by the students and media alike. From the problems shown above, the worst – as far as we, the authors, see it – is the lack of research done over this massive change in the school culture. The problems showed above are with reasonable assumption only a tip of an iceberg ready to collide the society. Thus the system should be analysed not only by the methods of an information ethicist but by a multidisciplinary group of scientists with long-time comparative empirical study. To emphasize the point: we do not know the effects of these changes.

The information processing, handling and most of all the upkeep must be handled – if not better – at least more openly in such way that the private information is ensured to stay at least between the adolescent and their parents (if even with the guardians). The gathering of the information should be standardized so that the teacher and the features etc. of the adolescent are not a factor in the generation of the information – at least not in a large scale.

Feedback given by the teachers should be something that helps adolescents to get better in school and in life, but as stated before teachers too tend to misuse the system by giving unconstructive feedback instead of following instructions of the system developer. This kind of feedback can by itself cause harm to adolescent receiving it. It is not a surprise that adolescent turn these blunt notions into humour since there is so little value to them other vice.

The teachers in public schools (overwhelming majority in Finland) use their mandate to exercise their government-related power over the citizens – the children and their guardians. Therefore there should be some mechanism to check whether the power is used fairly and within the limits of the laws and morals. Yet it seems the teachers can make notes to their pupils' school sheets without any reasonable risks of being accountable for their actions. Moreover they can justify harsher actions with the records they have written – true or false. This – at least according to the postings in Facebook – seems to be a thing the teens rebel against by doing illegal and immoral acts against teachers' person and possessions.

Yet the problem is not only between the adolescents and their teachers but moreover effects to the future of our whole society. When subjected to this kind of Panopticon-style information system it is possible that the children are accustomed to an electronic control both in the workplace and by the government. Do we really want to use a system which builds our children's' identity through fear and doubt rather than through cooperation and trust? If so, we will not raise a generation of citizens but a generation of subjects.

REFERENCES

[1] StarSoft (2015), *Wilman tuote-esittely* [What is Wilma], http://www.starsoft.fi/public/?q=fi/node/54

[2] Etelä-Suomen Sanomat (2011), *Lehtori: Wilma on yli-innokkaan opettajan ylläpitämä rikosrekisteri* [Lecturer: Wilma is a criminal record kept by an over-zealous teacher], Krista Koivisto / Etelä-Suomen Sanomat 27.2.2011, http://www.ess.fi/uutiset/kotimaa/2011/02/27/lehtori-wilma-on-yli-innokkaan-opettajan-yllapitama-rikosrekisteri

[3] Ilta-Sanomat (2013), *Suomen koululaisten uusi villitys huolestuttaa:"Lasten rikosrekisteri" leviää kaikkien nätäviksi* [Newist trend amongst Finnish schoolchildren rises concern:"Childrens criminal records" are spreading for everyone to see], Pauliina Jokinen / Ilta-Sanomat 19.4.2013, http://www.iltasanomat.fi/perhe/art-1288558599629.html

[4] Länsiväylä (2013), *Wilmasta on tehty lasten rikosrekisteri* [Wilma has been made to be a criminal record for the children], Jari Pietiläinen / Länsiväylä 16.4.2013, http://www.lansivayla.fi/artikkeli/232746-%E2%80%9Dwilmasta-on-tehty-lasten-rikosrekisteri%E2%80%9D

[5] YLE (2009), *Koululaiset: Wilma on rikosrekisteri* [Schoolchildren: Wilma is a criminal record], Kirsi Tirkkonen / YLE [Finnish National Broadcasting Company] 13.5.2009, http://yle.fi/uutiset/koululaiset_wilma_on_rikosrekisteri/5250971

[6] YLE (2014), *Wilma pilaa kodin ja koulun loputkin välit,* [Wilma ruins the remaining relations between home and school], Sari Helin / YLE [Finnish National Broadcasting Company] 4.2.2015, http://yle.fi/uutiset/sari_helin_wilma_pilaa_kodin_ja_koulun_loputkin_valit/7780204

[7] Erikson, E. H. (1982) Lapsuus ja yhteiskunta. 2. korjattu painos. [Childhood and Society] Jyväskylä: Gummerus.

[8] Furman, Ben (2013), *Viesti Wilmalla viisaasti.* [Communicate wisely with Wilma] StarSoft. 11.1.2015, http://www.starsoft.fi/public/?q=node%2F13298.

[9] Uusisuomi (2013), *Facebook-ilmiö oli liikaa: Wilma-viestit piiloon koululaisilta*. [The Facebook phenomenon was too much: Wilma messages hidden from schoolkids.] http://www.uusisuomi.fi/kotimaa/58704-facebook-ilmio-oli-liikaa-wilma-viestit-piiloon-koululaisilta.

[10] Nurmi, J-E. (1995*), Nuoruusiän kehitys: etsintää, valintoja ja noidankehiä.* In P. Lyytinen, M. Korkiakoski & H. Lyytinen (ed.) Näkökulmia kehityspsykologiaan. Helsinki: WSOY.

[11] Bentham, J. (1843) The Works of Jeremy Bentham, ed. Bowring, J., vol IV, available at: http://oll.libertyfund.org/titles/1925, accessed 16.2.2015.

[12] Foucault, M. (1991/1975) Discipline and Punish: The Birth of the Prison (translated by Alan Sheridan), Penguin Books, London, England.

From Participatory Design and Ontological Ethics, Towards an Approach to Constructive Ethics

Sandra Burri Gram-Hansen
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
+45 99407405
burri@hum.aau.dk

Thomas Ryberg
Aalborg University
Rendsburggade 14
9000 Aalborg, Denmark
+45 99407400
ryberg@hum.aau.dk

ABSTRACT
This paper explores, analyses and discusses the potential of applying Danish theologian and philosopher K.E. Løgstrup's ontological approach to ethics, when planning and conducting participatory design activities. By doing so, ethical considerations, will transform from being a summative evaluation perspective typically included at the end of a design process, to becoming a more formative and constructive perspective which influences the entire process. The approach presented in this paper will support on-going research within the field of Value Sensitive Design with theoretically based principles. These are principles that practitioners may consider when planning e.g. workshops in order to ensure that the activities facilitate both the design process and establish an ethical foundation for the design process. In addition to the theoretical contribution of the paper, the notion of constructive ethics is exemplified in practice by on-going research in the cross field between persuasive design and learning, carried out in collaboration with the Danish Military. Previous research has suggested that both participatory design and ethics may be essential to persuasive design in theory and in practice. However, considering the impact interactive technologies have on users in general, the principles exemplified through this case are relevant in a much broader perspective and to many other design traditions.

Categories and Subject Descriptors
H.5.3 [**Group and organizational interfaces**]: Computer supported cooperative work, Evaluation/methodology

General Terms
Design, Human Factors, Theory.

Keywords
Kairos, Løgstrup, Constructive Ethics, Participatory Design

1. INTRODUCTION
This paper explores and exemplifies how ethics may potentially be considered a constructive perspective in a design process, by including the ethical demand by K.E. Løgstrup when planning and executing participatory design activities. It is commonly acknowledged that ethics is an important perspective to consider in relation to technology design, even more so in areas where technologies are strategically used to influence the users' behaviour. However, ethical reflections are often included as a deconstructive evaluation process at the end of an otherwise dynamic and iterative design process. Whilst the arguments and exemplifications presented in this paper are primarily related to a particular human centred approach to persuasive design, the overall perspective, that ethics can and should be considered constructive to the design process, is both valid and applicable within all areas of HCI design. This paper provides an introduction to the fundamental aspects of Løgstrup's ontological approach to ethics, and exemplifies how this perspective can be taken into consideration in user centred design processes.

Several researchers have addressed the ethical challenges related to persuasive technologies, and exemplified how ethics may be taken into consideration in relation to intentional behaviour change [1-3] In 2009 Davis argued that Value Sensitive Design and Participatory Design might hold particular potential with regards to an ethical approach to persuasive design. In continuation of Davis' work, this paper argues that participatory design should perhaps be considered a requisite for persuasive design, and that by including the ethical reflections of Danish philosopher K.E. Løgstrup, the ethical perspective may be constructive for the design process. The aim of the paper is to provide a practical example as to how ethics may serve, as a constructive foundation for a design process, rather than be included only as an evaluative measure once major decisions regarding the design have already been made and potentially effectuated.

As research within the field of Persuasive Technology (PT) has progressed and developed, important perspectives relating to persuasion, PT and the potential of this particular approach to technology design, has been explored, exemplified and discussed. The notion of persuasive computers is most often explained with reference to Fogg's original framework, and particular interest has been directed towards the design principles introduced in the Functional Triad (FT) [4]. In continuation of Fogg's research, the term Persuasive Design (PD) has been widely applied in a variety of contexts, but without a common definition, making it challenging for researchers to clearly pinpoint the unique claim of PD when it is applied to more established research areas [3]. The design principles do not constitute novel approaches to technology design, but are the result of extensive research in understanding the persuasive potential of interactive computer technologies. As a result, the principles are already applied within many established

design traditions, and research shows that within fields such as Information Architecture, Digital Dissemination of Cultural Heritage and Technology enhanced Learning, they do not lead to new design insights [2, 5]. As a result, previous research has suggested that a clear distinction between PT and PD is necessary in order to establish the claim of PD, when applying the concept and principles to other more established research areas.

PT is generally understood as technologies that are designed with the intent to change attitudes and/or behaviours. Often given examples of areas where PT are applied include exercise motivation, healthcare, environment preservation and energy saving. However, this definition is also subject to some discussion as all design can be seen as inherently persuasive and designed with the intent to motivate particular user behaviour [4]. In our approach, PT is understood as a technological focus, with emphasis on usability, system design and user feedback.

PD on the other hand, may be more easily understood if it is seen as a wider and more nuanced approach to design, where understanding and incorporating contextual reflections in the design process, is considered key to ensuring the persuasiveness of the design [2, 6]. The understanding of context and appropriate situation is primarily understood in relation to the rhetorical notion of Kairos. Through this approach Persuasive Design can be considered a meta-perspective which may be applied to more established design traditions and which emphasises ethical reflections in the design process and a particular focus on the intended use context. Furthermore it is argued that the claim of PD may not be constituted by the specific technology alone but on the adaptation between the persuasive technology and the intended use context. As such, this approach to PD acknowledges the theoretical and methodological perspectives within the PT field (e.g. Fogg's argument that technologies hold a particular persuasive potential), but distinguishes between PD and PT in order to clarify how PD may be a benefit to more established fields of technology design. Whilst the design principles discussed and applied in PT are also seen used in other research fields, the context oriented PD perspective justifies how the principles become persuasive when applied in the intended use context.

In continuation, learning has been argued to constitute a foundation for persuasive design, based on the understanding that a distinction between nudging and persuasion, may be found in the persuasive aim to achieve sustainable attitude and behaviour change. In order for the behaviour change to be sustainable, it must be based on an attitude change, and in order for a person to change attitude towards a topic, he or she must acquire and process new knowledge. – Or in other words learn [7]

2. KAIROS – A MATTER OF APPROPRIATENESS AND CONTEXT ADAPTATION

The mentioned distinction between PD and PT is primarily based on reflections regarding the rhetorical notion of Kairos. The link between PT and classical rhetoric was brought to attention by Fogg, and as the theoretical foundation of PT has been explored and further developed, some researchers approach the challenges of this novel field from a foundation in classical humanistic traditions that include rhetoric, logic and ethics [8]. Kairos is often referred to as an essential in relation to persuasion, and is most often described as the opportune moment to act or trigger a persuadee into changing attitudes or behaviour [9]. The concept sums up the principle that any rhetorical approach is based upon the specific situation, and that comprehension of the context as such is one of the most vital resources when deciding upon rhetorical means to apply to a given argument [10, 11]. Hansen specifies that the definitions of Kairos vary from narrow translations such as "particular point in time" and "specific circumstance", to wider concepts such as "situation", "occasion", and "opportunity". The narrow and wider definitions of Kairos are inseparable and must be considered in relation to each other.

Kairos is three-dimensional and comprises the appropriate time, place, and manner to address the persuadee. As was the case with Hansen's distinction between a narrow and a wider definition of Kairos, the three dimensions of Kairos are also inseparable and must all be equally taken into consideration if the opportune moment is to be defined. Nonetheless, Kairos is most often referred to in relation to timing, or in continuation, the ability to act at the right time and in the right place. Reflections regarding the appropriate manner are seldom given the same considerations. Nevertheless, when acknowledging all three dimensions of the concept, Kairos constitutes not only an opportune moment, but also an understanding of what is appropriate within a given situation.

With the term *appropriate*, Kairos emphasises the importance of the performed action being both effective and also ethical. Although ethical debates often discuss the difference between right and wrong in a given situation, ethics per definition may just as well focus on the appropriateness within a given situation [14]. The perception that persuasion must take place in an appropriate manner, does not only refer specifically to the design of a system, but also to a general understanding of the context in which the technology is to be applied and it is this wider contextual understanding of Kairos which may be directly linked to ethical considerations. In light of the challenges related to defining the claim of PD in relation to more established fields, Gram-Hansen et.al. argue that the strong demand for ethical reflections in relation to PD, may in fact be one of the aspects which specifies the relevance and claim of PD when applied in well established areas of application such as learning, information architecture and digital mediation of cultural heritage. While persuasion in other design traditions is acknowledged as an integral part of a design process, PD enriches the design process with a communicative determination and a demand for a recurring ethical evaluation process [2].

Besides from defining ethics as a key concept in PD, the multi-layered definition of Kairos and in particular their inseparability, also gives reason to consider the relationship between PT and PD as multi-layered and inseparable. PD may be considered a particular type of context adaptation, which focuses on establishing the appropriate balance between the technology and the intended use context. Many of the technologies, which dominate the current landscape of technologies, in particular context-aware devices such as smartphones and tablet computers, can be argued to hold the potential to meet both the temporal and location-based dimension of Kairos. However the persuasiveness of each technology is dependent on the device also being applied in the appropriate manner within the given context. This third and final dimension calls for a different and more nuanced evaluation of the intended use context, it may not be formalized, and it may be argued to point towards the necessity of not only designing the appropriate technology, but also design the appropriate balance between the technology and the intended use context [6].

In practice, this approach to PD distinguishes itself from PT, in a way that acknowledges the theoretical and practical steps taken so far within the PT research community, and facilitates the

persuasive potential of the technologies by adding a wider and more context oriented layer. System- oriented methods such as the PSD model [12], address the challenges related to the specific technology design, whereas PD is considered a wider concept which focuses on the establishment of an appropriate balance between technology and context, and which may serve as a meta-perspective to more established research fields. Most importantly however, the notion of a multi-layered approach to design is brought into consideration due to the interdependency between PT and PD.

3. PARTICIPATORY DESIGN AND AN APPROACH TO CONSTRUCTIVE ETHICS

As mentioned in the previous section, the notion of Kairos is already widely acknowledged as an important perspective to consider in relation to persuasion.

It has previously been argued that value sensitive design and participatory design may hold the potential to incorporate ethics in the design process when developing PT [1]. When considering Gram-Hansen's perspectives on Kairos and CA, it may be argued that Participatory Design is not merely something that can be considered but something that must be considered if a persuasive design is to incorporate all three dimensions of Kairos. There are numerous methods for exploring intended use contexts and gaining a better understanding of the activities, which commonly take place. However, Participatory Design distinguishes itself by designing technologies with user participation, and by focusing on the users' understanding of that context and in continuation the users' understanding of what may and may not be appropriate. As such, if PD is to incorporate all three dimensions of Kairos, Participatory design becomes a prerequisite to the overall design process.

Besides from providing important user insight, participatory design is also argued to hold the potential to overcome some of the ethical problems related to persuasive technology [1]. As mentioned, researchers have argued that a participatory design approach may be a way to address this challenge. However, it is important to acknowledge that simply involving the users in the design process is insufficient to ensure that the practical and ethical results of a participatory design activity is carried through to the subsequent steps of the design process. Participatory design offers a range of methods and reflections regarding user involvement, and workshops that include games, role-playing, or inspiration cards are being widely applied in a variety of design fields. However, the challenge remains that many of the results that are reached through participatory design may be difficult to conceptualize and fully implement in the final design of a technology. This challenge relates to both the practical design input and suggestions made by the user participants, but also the ethical aspects, which may have been touched upon. In order to address this challenge we argue that ethics should be incorporated as a constructive foundation for the participatory design process, and that this may be done by considering the reflections of Danish philosopher K.E. Løgstrup (1905-1981).

4. LØGSTRUPS ETHICAL DEMAND

Løgstrup was a Danish philosopher and theologian who has manifested himself as one of the great Danish thinkers. He presented his approach to ethics as based on the so-called ontological tradition. According to this tradition, humans are influenced by basic conditions that are inalterable. For instance, the life of a human is inevitably entangled with other humans from the very moment we are born, and any type of human interaction results in a relation of ethical significance. Thereby, Løgstrup's approach to ethics distances itself from the both the utilitarian and the deontological tradition, by rejecting the possibility of evaluating ethics objectively (based on either actions or the consequences of such), and emphasising that ethics must be considered intuitive and open to be influenced by all humans.

Løgstrup argues that humans are born with several characteristic referred to as *the sovereign expressions of life* which include features such as benevolence, compassion, trust, love and open speech, and that these qualities are essential for the interaction between human beings. Caring for other humans is simply part of human nature, or as he calls it, *the ethical demand*. The spontaneous manifestations of life can as such be considered the features within human nature which are generally viewed as ethical, contrary to characteristics such as jealousy, hate, mistrust and injustice.

"The demand, precisely because it is unspoken, is radical. This is true even though the thing to be done in any particular situation may be very insignificant. Why is this? Because the person confronted by the unspoken demand must him or herself determine how he or she is to take care of the other person's life."

(Løgstrup 1997, 44)

The ethical demand in itself is silent; in the way that Løgstrup does not attempt to set up rules concerning ethical and unethical actions. Contrarily, Løgstrup argues that the individual performing the action, in accordance with the reality perception of that individual must make the assessment of the ethicality of actions taken in a given situation. Humans must be conscious that any type of human interaction results in a situation where one human becomes responsible for the life of another human being and in accordance with such acknowledgement; humans must strive towards doing to others as they trust others to do to them [13].

By defining ethics as an intuitive result of human nature, rather than moral rule based on reason, Løgstrup opposes one of the most recognized philosophers of deontological ethics; Immanuel Kant, who is known especially for introducing the categorical imperative, which promotes the idea that ethics is a matter of acting rationally. Løgstrup makes the argument that ethics based on the human ability to think freely is problematic, as this ability also enables the human mind to justify an action that at first hand does not appear ethical at all. Løgstrup states that humans in general have a clear sense of what is right and what is wrong, but that they also tend to end up in situations where conflict arises between the ethical choice and obligations bound in for instance legislation or profession. Police officers may find themselves arresting citizens who are breaking the law, but who may be doing so for reasons that could be considered ethical e.g. stealing in order to feed a starving family. In that case, the police officers may find themselves acting against their ethical demand and justifying it by referring to the requirements of their job. In such situations, humans tend to excuse acting against their ethical duty to an extent where the excuses themselves end up appearing as committing as the original ethical duty. The result is a balance between the ethical and the obligated action, which allows the human to choose freely between the two, and thus acting against the ethical duty [14]

The intuitive nature of Løgstrup's ethical perspective and the silent claim of the ethical demand make it difficult to apply when ethically evaluating a situation or a technology. However, Løgstrup's reflections regarding human interaction and the sovereign expressions of life can be considered when planning

and preparing participatory design activities. Løgstrup's ethical reflections may direct the facilitator of a participatory design activity to strongly consider ways to ensure a mutual power balance between the participants of for instance a workshop, in order to ensure that all participants feel that they are free to speak their mind, and to ensure that a mutual responsibility is established between the participants. By reference to Løgstrup, we will argue that if a mutual understanding is established between the participants, they will be more motivated to consider the results of a workshop in later work, and they will intuitively aim towards meeting the ethical requirements of the other participants. As such, the ethical perspective may become constructive for the design process, rather than be included, as a an evaluative measure that is applied either during or after the design process has been finalized.

5. FROM WARRIORS TO CLIMATE WARRIORS – PROJECT GREEN BARRACKS

In the following section, the approach to participatory design and constructive ethics is exemplified by a workshop held with participants from the Danish Military Defence in September 2013. The workshop exemplifies specific areas in which Løgstrup's ethical perspective may facilitate the planning and execution of a workshop. Naturally however similar reflections and adjustments are required in other steps of the design process.

The energy and environment related challenges that the world is facing are well known to most. The Danish Military Defence is one of the largest organisations in Denmark and as a result the Danish Ministry of Defence has presented an ambitious climate and energy strategy, which addresses ways in which the Danish Military Defence actively wishes to lower the energy consumption level and minimize their influence on the climate.

An extensive deal of the collective energy consumption in the Danish Military Defence is related to the use and maintenance of buildings and military establishments. This has led to the large scale Project Green Barracks in which the Danish Military Defence in collaboration with industrial partners, educational institutions and other innovative partners, are working towards innovative and rational solutions that will reduce the energy consumption. Two existing military establishments have been selected for pilot studies, namely Aalborg Kaserne in Northern Jutland and Almegårds Kaserne on Bornholm.

Amongst the challenges related to the project is the attitude of the employees who work in the Danish Military defence. Although most will agree that climate and energy consumption are challenges that must be considered, also by the military defence, many find it hard to relate the solutions to their current work processes. For the soldiers who train for missions abroad it makes little sense to worry about switching of the light when leaving a barrack to rush out on a dangerous mission, and most find that while the focus of their work is on military duties, the aim of saving energy and the solutions for reaching that goal must be dealt with by the establishment administration office. This particular attitude collides with the overall strategy of moving towards "Green Establishments" where energy saving is a fully integrated element in the work procedures of the organisation. As such, the project aims not only to physically develop Green establishments for the future, it also focuses on a change of both attitude and work practice of the employees.

One of the initiatives taken within Project Green Barracks is a competition for architects to deliver the best solution for future military defence establishments. The competing architects were to present solutions, which are not only sustainable and innovative, but also consider the work practices and the work environment for the organisation employees and collaborators. In order to provide the competitors with sufficient and suitable information about the work practice a series of workshops were held by the establishment administration office, enabling employees from all areas of the organisation and different ranks to contribute with knowledge and suggestions. In the following one specific workshop is described as an exemplification of how participatory design activities may be based on a constructive approach to ethics.

6. DESIGNING FOR THE FUTURE

In September 2013, 18 representatives from the Danish Military Defence met to engage in defining requirements for the future green establishments. The participants included representatives from the Danish Ministry of Defence, The Danish Defence Installation Command and different level 3 authorities (high authority members of the practical and educational staff) from the two pilot establishment included in Project Green Barracks. The group of participants were all sufficiently experienced within the military defence to be able to share insightful knowledge about current practices. Most of them had also served on missions abroad, and most of them will still be active within the military defence in 20 years and will as such be part of the transition that the organisation is expected to go through. The specific aim of the workshop was to collectively define visions and requirements for the future military establishments, but without suggesting solutions. For instance, it was accepted for the participants to state that it is a requirement that the military employees can train all year round regardless of weather conditions, but it was up to the architects to decide how that requirement can be met.

Whilst the workshop was being planned and prepared, Løgstrup's ontological approach to ethics was taken into consideration as a constructive framework, leading the designer of the workshop (first authour) to reflect carefully on ways in which the physical location and the different activities might influence the power balance between the participants. The overall goal was to establish a workshop that would motivate the participants to engage and discuss possibilities for the future, and also to establish a mutual understanding between the participants.

The Danish Military Defence considers itself a particular type of educational institution with a specific educational profile and with a history of a strong hierarchical structure amongst the different employees. The strong sense of hierarchy is common within military defences world wide, and can be argued to be vitally important when those serving are posted in critical areas abroad. However, when planning a workshop where everyone's opinion is considered equally important, this natural power balance may be a challenge. As a result, Løgstrup's perspective on interaction, power balances and the sovereign expressions of life were carefully considered in relation to both the location and the different workshop activities. Løgstrup opposes deontological philosophers such as Immanuel Kant, by arguing that ethics, rather than being based on reason, is based on the human ability to think freely and act on intuition [15]. Furthermore, ethics is argued to spring from interaction, as humans are inevitably entangled and influence each other through the way we interact.

"Through the trust which a person shows or asks of another person, he or she surrenders something of his or her life to that person. Therefore, our existence demands of us that we protect the life of the person who has placed his or her trust in us. How much

or how little is at stake for the person who has thus placed his or her trus in another person obviously varies greatly"

(Løgstrup, 1997:17) [13]

6.1 Workshop Location

Considerations regarding ways in which the physical location might influence the interaction and communication during the workshop, lead to the decision to host the workshop at Aalborg university's E-Learning Lab. Originally the workshop had been scheduled to take place at a military establishment in central Denmark, but this would potentially have allowed some participants to be on their "home turf", and as such feel more secure and potentially be more dominating. By moving the location of the workshop to the facilities at Aalborg University, all participants found themselves on uncommon but neutral ground and in that sense equally vulnerable. It did however also constitute a context, which motivated new reflections. It was emphasised that the workshop was to be considered a "safe area" in which everyone was encouraged to speak freely. The rationale of the designer was, that if the workshop had been held at a military establishment, the participants could potentially be influenced more by the context and find themselves less able to rise above their well known behavioural patterns.

6.2 Workshop Activities

In practice the workshop was facilitated as a full day schedule starting at 9am and ending at 3.30pm. The overall structure of the workshop was inspired by Jungk and Müllers approach to future workshops in which participants engage in a critique phase, and an utopian phase and a solution phase [16]. In consideration of the mentioned requirements for this particular workshop the method was however adjusted and consisted of only two phases;

- The Critique phase
- The Future phase.

The solution phase was not included in workshop, as the aim was to provide the competing architects with visions and inspiration regarding future green army barracks, rather than providing them with solutions or specific requirements.

During the critique phase, participants were presented with reflection exercises that would motivate them to reflect upon their own work practices and the challenges that they come by in their daily work. Even though the participants share educational backgrounds and general work tasks, there are important distinctions between their daily work practices. Not all military establishments are identical in Denmark and depending on the primary tasks which a carried out at a given establishments, the maintenance of the buildings may have been directed towards different areas. In order for the participants to collectively discuss visions for the future, they first had to establish a mutual understanding of current challenges and in some ways a mutual language to discuss these challenges.

In practice, the critique phase consisted of three different assignments which all aimed at motivating the participants to reflect on current challenges, and to explain these challenges to the rest of the group. In consideration of the aforementioned hierarchical structures and the existing power balances, the first two assignments were individual and aimed for very specific exemplifications of current challenges. The individual tasks were planned to ensure that all participants were provided with an opportunity to air opinions, prior to engaging in group work. The third assignment was solved in small groups of six participants and aimed for a shared understanding of the general challenges of the current military establishments. The critique phase was concluded by a group discussion of the aspects, which were perceived as particularly challenging.

In terms of ethical considerations, the critique phase served as an icebreaker for the collaboration in general, as well as a gentle initiation of establishing mutual understanding and trust. The practical assignments required very little effort from the participants (they were given the option between drawing and writing and asked to do whatever made them most comfortable), and they were asked to simply consider their own everyday work practice and explain challenges. By addressing different work related frustrations through a specific workshop task, the participants were provided with a safe space to express their opinions freely – contrary to what might be appropriate in their normal work environment. In continuation, the task uncovered mutual frustrations amongst the participants, which helped establish a sense of community in spite of different work locations and everyday responsibilities.

Whilst the critique phase uncovered mutual understandings of the current challenges, the subsequent Future phase focused on defining the participants' requirements, needs and visions for the future military establishments.

All assignments in the future phase were solved by the previously established groups, and the results of the critique phase were brought into consideration as a common starting point for this second phase of the workshop. Furthermore, the participants were asked to also consider a list of trigger words defined by the Project group behind Project Green Barracks.

Image 1 -The future phase began with production of red and yellow concept cards

In order to maintain the collaboration and discussions of the previous phase, the first assignment was development of concept cards. The participants were provided with red and yellow cardboard pieces and asked to create as many cards as possible by letting red cards represent actions which should take place in the future establishments and yellow cards represent facilities which would be required to do the actions. The goal of the activity was to motivate the participants to consider the requirements of their everyday work activities and to break these requirements into specific concepts, which might serve as an inspiration to the other groups. Also, the cards were used as a tool to spark conversations amongst the participants about what activities really need to take place in the army barracks, and to provide the designs with rich explanations as to why certain practices take place. For instance, outsiders may consider the old-fashioned bunk beds and small closets out-dated, but through the card activities it was explained

that they actually serve an educational purpose. If the soldiers are unable to adjust to having no privacy and very little space when they are safe and at home in Denmark, they will most certainly not be able to work and live under similar conditions when dispatched to war zones around the world.

All cards were placed on pin boards and all participants were encouraged to read them and use them as inspiration as they went on to prototyping.

Image 2 - Solutions were discussed as prototypes of the future Danish army barracks were developed with Crayons, Lego and Play dough

The primary assignment of the future phase was the development of prototypes of the future green military establishments. Each group was provided with a large green surface, crayons, play dough, and Lego, and told to create the facilities, which they found essential to the work practices within the military defence. The groups were given completely free hands with their prototypes; however they were required to specifically explain what the intent was behind each facility in the prototype and explain its overall relevance for the future green military establishments.

The artefacts applied in this face, were carefully chosen partly to ensure that the produced prototypes remained conceptual and not too specific, and partly to maintain the mutual understanding and balanced hierarchical structure, which had been established through the previous activities and facilitated by the location.

Very often, workshops include asking the participants to draw something. It is considered an easy task, as most people have been used to drawing since childhood. However, far from everyone feels confident drawing and sharing their creations with others, and as a result, some participants may be reluctant to share what they produce and explain its meaning.

Contrary, Lego and Play dough are artefacts that leave little or no room for detail, and they hold the benefit that most people are equally skilled in using them for production. Often, only the participants involved in building something with these artefacts, will know specifically what the result represents (a building, a vehicle etc.). As a result, Lego and Play dough constitutes artefacts that not only ensure that prototypes remain conceptual, they also help to ensure that participants do not back out of the design phase due to feeling incompetent.

The future phase was concluded by each group presenting their prototype to the remaining participants, and finally a round discussion about the overall outcome of the day.

7. CONCLUDING REFLECTIONS

In this paper it has been argued that ethical reflections may be a key element in explaining the unique claim of PD, and furthermore that the multi-layered notion of Kairos gives reason to consider PD a particular type of context adaptation. PD may be considered a meta-perspective, which focuses on establishing the appropriate balance between the PT and the intended use context, thus enabling the technology to reach its full persuasive potential. In order to do so, participatory design also becomes a requisite for PD, as only the intended users have the insight to express what may or may not be considered appropriate within a given context.

Participatory design is already an established and often applied approach to technology design, and it is becoming generally acknowledged that in order to develop technologies that will be accepted and applied by the users, we must include the users in the design process. The particular distinction with the described example of participatory design in the Danish Military defense lies in the foundation of the workshop, which is based on a constructive approach to ethics. By considering Løgstrup's ontological approach to ethics during the preparation of the workshop, the workshop served not only as a way to involve the users in the design process, it also helped establish a potential ethical foundation for the future development process.

In particular, Løgstrup's ethical perspective lead to reflections regarding:

- **The power structure between workshop participants** (How could the design process influence the existing power structure in a way with allowed all participants to feel empowered and safe to express their opinions)
- **The Location** (How would the location influence the workshop activities and the previously mentioned power balance)
- **The workshop room** (How did the room influence the interaction between the users, and how could the physical surroundings help adjust the power balance between participants)
- **The workshop activities** (Which activities would enable and motivate balanced interaction between the workshop participants, and still lead to the required end results)

As the overall aim of the workshop was to generate user insight for the architects to take into consideration as they progress with their competition bids, the actual prototypes developed in the workshop were not the primary outcome of the day. The presentations and discussions that took place on the other hand provided much valuable insight to the participants understanding of their work context. As the participants were asked to present their reflections after every single assignment throughout the workshop they were constantly required to reflect and explain their position in detail. The prototypes served the important purpose of specific exemplification, and did as such motivate the participants to be very precise in their descriptions. Through the interactions that have taken place during the workshop, and with the establishment of a mutual understanding amongst the workshop participants, an ethical foundation may have been constituted. Løgstrup argues that humans are ontologically bound to each other, and as a result, the ethical foundation established through the workshop holds the potential to also become an influencing factor in the following development process as well as in other initiatives taken within the Project Green Barracks.

Løgstrup furthermore argues that once we engage in interaction and influence each other's lives, our experiences shape us and affect the way we subsequently interact with others. When we understand the challenges our users are facing we are more likely to give them the consideration they demand. The establishment of an ethical relation between workshop participants is vital to the success of the Green Establishments, as the participants represent the military employees who will responsible for carrying out green initiatives in practice. Empowering the participants and providing them with a safe context for sharing ideas and reflections, is expected to constitute a motivating factor once the participants return to their normal work areas and share their newly gained experience with their colleagues who did not participate in the workshop. As such, the workshop not only provides insight regarding the participants understanding of appropriateness within their work context, it also serves as an initiator of the ontological framework which forms the basis of Løgstrup's approach to ethics.

The overall benefit of applying Løgstrup's ethical reflections when planning a design process, does not only apply to the case described in this paper, it can be considered generally applicable. By allowing ethics to form a constructive perspective in the design process, and for instance include reflections regarding ways in which physical surroundings and different activities influence the power balance between those partaking in the design process, ethics becomes fully integrated throughout the design process.

Different activities whether they be traditional meetings, interviews or participatory design oriented, provide different insights to the requirements of a design, but the understanding of the requirements and the motivation to meet them are to some extent dependent on a mutual understanding which can only be created through interaction between participants who are equally empowered.

Due to the intuitive nature of Løgstrup's approach to ethics, it must be emphasised that the approach presented in this paper should not replace other traditional approaches to ethical evaluations but rather be seen as a supplement. When considering intentionality as well as the nature of persuasion as being focused on attitude and behaviour change, both the deontological and utilitarian approach to ethics still constitutes essential perspectives. We do however find that by considering Løgstrup's approach to ethics in relation to participatory design, we are able to argue towards a constructive approach to ethics in the design process, which facilitates the requirements of ethical reflections in PD, and at the same time enables designers to fully reflect upon all three dimensions of Kairos in the development process

8. ACKNOWLEDGEMENTS

The workshop presented in this paper was arranged in collaboration with Boie Skov Frederiksen, Thomas Troels Klingemann and Thilde Møller Larsen from The Defence Installation Management Command, and with the support and collaboration of Claus Uttrup, Brigadier general, Chief of Staff, Installation Management Command.

9. REFERENCES

[1] Davis, J., Design methods for ethical persuasive computing, in Proceedings of the 4th International Conference on Persuasive Technology. 2009, ACM: Claremont, California.

[2] Gram-Hansen, S.B.a.L.B.G.-H., On the role of ethics in Persuasive Design, in Ethicomp 2013. 2013, Syddansk Universitetsforlag: Kolding. p. 8.

[3] Redström, J., Persuasive Design: Fringes and Foundations, in Persuasive Technology 2006. 2006, Springer: Eindhoven.

[4] Fogg, B., Persuasive Technology, Using Computers to change what we Think and Do. 2003: Morgan Kaufmann Publishers.

[5] Lykke, M., Persuasive design strategies: means to improve the use of information organisation and search features in web site information architechture?, in ASIST Special Interest Group on Classification Research 20th Workshop. 2009: Vancouver.

[6] Gram-Hansen, S.B. and T. Ryberg, Persuasion, Learning and Context Adaptation. Special Issue of the International Journal on Conceptual Structures and Smart Applications, 2013.

[7] Gram-Hansen, S. and T. Ryberg, Acttention – Influencing Communities of Practice with Persuasive Learning Designs, in 10th International Conference, PERSUASIVE 2015, T. Mactavish and S. Basapur, Editors. 2015, Springer: Chicago. p. 184-195.

[8] Hasle, p.a.A.-K.K.C. Classical Rhetoric and a Limit to Persuasion. in Persuasive Technology. 2007. Palo Alto: Springer.

[9] Aagaard, M.a.P., Øhrstrøm and Lars, Moltsen. It might be Kairos. in Persuasive 08. 2008. Oulu Finland: Springer.

[10] 10. Hansen, J.B., Den rette tale på det rette tidspunkt. RetorikMagasinet, 2009. **74**.

[11] Kinneavy, J.L., Kairos in Classical and Modern Rhetorical Theory, in Rhetoric and Kairos, Essays in History, Theory and Practice, P. Sipiora and J.S. Baumlin, Editors. 2002, State University of New York Press.

[12] Oinas-Kukkonen, H.a.M.H. A systematic Framework for Designing and Evaluating Persuasive Systems. in Persuasive 2008. 2008. Finland: Springer.

[13] Løgstrup, K.E., The Ethical Demand. 1997: University of Notre Dame Press.

[14] Muckadell, C.S.d., Løgstrups Etik. En moralfilosofisk blindgyde. 1997: Gyldendal.

[15] Gram-Hansen, S.B. Towards an Approach to Ethics and HCI Development, based on Løgstrup's Ideas. in Interact. 2009. Uppsala: Springer.

[16] Jungk, R.a.N.M., Future Workshops: How to Create Desirable Futures. 1987, London: Institute for Social Interventions.

Between Insanity and Love

Ryoko Asai
Uppsala University
Box337, 75105
Uppsala, Sweden
ryoko.asai@it.uu.se

ABSTRACT

Technology has opened up more opportunities to find better partners, especially via online dating sites . In addition, technology related to love and sex currently goes far beyond online dating sites. Technology influences our intimate life more and more. Media started to pick up romantic relationship between human beings and digital characters often. Furthermore, today, many wearable devices to experience virtual sex have come into the market. And also robots designed for having a sex with human beings are being developed rapidly. Some might claim having a sex without love or without reproduction is just totally pointless. Or, having a sex with robots is totally "insane". But apparently there are big market needs, and modern technology seems to be able to satisfy them. This paper explores how it is possible for us to feeling love or sexual desire for non-organic objects by conducting the interview survey, and also considers why people want to have technology for satisfying sexual desire from a philosophical perspective.

Categories and Subject Descriptors

K.4 [**Computers and Society**]: Ethics; K.8.0 [**Personal Computing**]: General—*Games*

General Terms

Theory

Keywords

Eroticism, insanity, love, sex, technology

1. INTRODUCTION

How did/do you find your partner? Information and communication technology (ICT) generated the new place for both men and women to find partners. There are many, almost countless online matching/dating sites in the world. Many online dating companies offer various services to users regardless of users locations, living time zones, nationalities and so on. As long as users can communicate with others who might be a future partner, they can find someone living in foreign countries or even on the other side of the world. People enjoy lots of increasing opportunities to meet "someone who might be for me" online. For example, in Sweden, meeting and finding partners online is the most popular than meeting in person in"meatspace".[1] Some news media report it as a hot topic or a new stream of finding partners on their magazines or TV programs. [2]

Normally people think if they have more options they could choose better one of those than choose one of few options. In the online dating, people have a much better chance of meeting a perfect partner beyond the national borders and cultural differences. However, too many choices require people more endeavor to "balance the tradeoffs between accuracy and effort" based on their heuristic strategies[1]. By contrast, the limited number of choices lead people to rational optimization. On the other hand, a number of options make the decision making process harder and also decrease people's satisfaction of the consequences of their choice. In other words, people exert more effort to take a better and more preferable choice among a huge numbers of choices and expect to get the better consequence.

Then, what happens in the end? The high expectation bring them less satisfaction, or sometimes it might let them down or regret. Because when they get the consequence regardless of positive or negative ones, they consider about other choices which they didn't choose [2]. Moreover, even though they believe they could choose the better partner whose personality is much more similar to themselves than choose someone in off-line occasion, the similarity of personality doesn't increase their happiness and satisfaction so mcuh [3][4][5]. But still ICT and technology created great communication channel in finding partners and make it possible to bring together one and the other who have little opportunity to meet persons of finding marriage partners for those whose social circles are rather limited. Additionally advanced technology generated new dimension of *love* relationship.

[1]SvD NYHETER, "Var fjärde relation börjar på nätet", 2010/4/15. According to the resent research by TNS SIFO, among the 1,111 respondents (25-60 years old), 23 percent of them met partners first time online as the highest percentage. The second popular way to meet partners is "meeting through friends and acquaintances (21percent)".

[2]For example, it is called as "Supermarket Love" in The Economist. The Economist, "Sex and love: The modern matchmakers", 2012/2/11.

2. MEETING A NON-ORGANIC PARTNER

Advanced technologies opened up interesting possibilities for a non-organic existence, such as two-dimensional characters, sex dolls or robots, to become a partner in *love* relationships. This study attempts to reconsider what love is in the highly technological society, and also to explore how to be ethical in incorporating technology in sexual activities. Generally we think *love* is one of fundamental emotions for animals, especially for human beings. However, it is difficult to explain what love is clearly and precisely. Aristotle's *Nichomachean Ethics* says that *love* is possible only between human beings [6] [7]. If it is possible to recognize *love* relationship only among human beings, how do we explain "love" for pet animals?

2.1 Rational use of technology in *Love*?

In social science, love is sometimes considered as a social exchange process [8]. It explains well why one-sided love is difficult for the long run. To love someone for long term, the relationship requires couples to make mutual efforts to keep love emotion. Once a person notices his/her love remains unrequited, it might be difficult for him/her to keep love in heart. That means absence of social exchange process in a relationship. In other words, reciprocity or mutual feeling is necessary for us to maintain love. Therefore, many thoughts and theories basically premise the relationship based on love establishes between human beings, or at least between living animals. Many theories, especially in economic theories, suppose that any kinds of action has any purpose (a particular utility) to be done by the action. In that sense, people would choose the rational way to achieve the purpose and to maximize utility [9]. Seeing sexual activity as a purposeful action, the main purpose which we come up with easily would be to reproduce next life.

It explains well why many people register their profiles on the matching sites and try to pick a promising "perfect" partner among huge numbers of candidates around the world. And under the situation where a number of choices are offered, action of trying to find a better partner online could satisfy a wide variety of needs and sexual preferences easier and more efficient than in the off-line occasion. Online matching sites require users to answer many questions about private information including sexual tendencies and their wish to have a child or to be a have-not. They can choose better ones depending on others' profile information among huge candidates. It seems to be very rational to use online matching sites in terms of finding a partner efficiently. Adopting technology in constructing the relationship contributes to generate more meeting places for people [3].

2.2 Loving something rationally and purposefully?

Furthermore, at the present time, technology is used not only for creating meeting places as described above but also for stimulating people's feeling of love and sexual desire. Technology has developed enough to interact with human beings and many interactive tools have permeated into society. For example, Sony had designed and marketed pet robots *AIBO* from 1999 to 2006 (see Figure 1 [4]) [5] . AIBO was designed like a puppy and it can mimic puppy's actions and original movements by programming. It cannot work as factory robots and it does never help people do household. The main purpose of it is only for fun on daily life and to be a pet of people. *AIBO* became a huge hit globally and people got it as their immortal pet. Many of them provide empathetic care to the pet robot and treat it as a real dog or a living animal. However, Sonny stopped aftercare and maintenance services for *AIBO* in the end of March 2014. The news of stopping maintenance services by Sony upset the owners and brought them sadness. Stopping maintenance meant AIBO is not immortal anymore. The owners scrambled its body parts and sought for a craftsperson who could repair their pets. News media reported many owners emphassized their love to their pet robots and some of them brought their broken pets to the temple for a funeral [6]. Following the conventional *love* definition, AIBO owners' feeling to their own robot pets is hard to see it as *love*. Love is generated between at least living animals and it is supposed to be evoked in the reciprocal relationship.

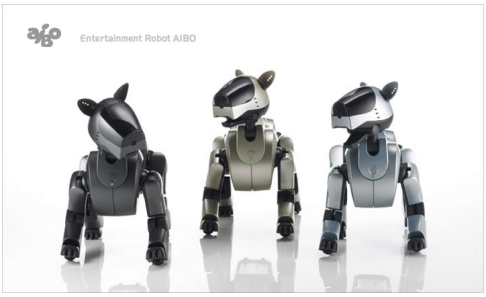


Figure 1: Pet Robot *AIBO*

When people have special feelings to non-organic existence, non-organic one doesn't necessary exist materially. There are many people who have special empathy or 'love' for two-dimensional characters such as gaming characters or animation characters. In the white paper on measures on declining birthrate issued by Japanese government, 37.6 percent of 20's and 30's unmarried men and women don't want to have a partner in Japan. The reasons why they don't want to have partner are: Don't want to be bothered to have the relationship (37.6 percent): Want to focus on the own hobbies (45.1 percent): Want to focus on studying/working (32.9 percent): No interest with love (28 percent) [10]. Earlier the white paper, a BBC documentary program covered the Japanese low fertility problem and picked up two Japanese men who prefer to have sex with virtual gaming characters [7]. Two Japanese men appeared in the BBC program, and they said they love their favorite gaming characters enough to satisfy their sexual desires [8] If it is really possible for

[3]Needless to say, using Online matching sites is not only for finding *love*but also for finding good friends, an one-night stand lover and so on. However, generally speaking, finding a partner for love relationship is most common there.

[4]http://www.sony.jp/products/Consumer/aibo/

[5]AIBO is an abbreviation of 'A'rtificial 'I'ntelligent ro'BO't.

[6]For example, *The New York Times* reported AIBO funeral in the temple and AIBO owners interviews on 17th June 2015. See more: http://www.nytimes.com/video/technology/100000003746796/the-family-dog.html

[7]BBC Two,"No Sex Please, We're Japanes" 3rd December 2013. You can see more information about this program online: http://www.bbc.co.uk/programmes/b03fh0bg

[8]In 2013, Japan's total fertility rate was 1.43. This rate is

us to love two-dimensional characters which has no physical existence and no emotions, there is no reciprocity and emotional exchange between human beings and two-dimensional characters. And what is the rational reason to 'love' two-dimensional characters?

3. REALITY OF *LOVE* FOR TWO-DIMENSIONAL CHARACTERS

In order to get to know how ICT and technology influence people's feeling of love, this study conducted the interview survey with six people who have their own favorite digital characters or two-dimensional characters as some kind of virtual partners. In Japan, it is well known globally that there are many ardent anime fans those are so-called *otaku*[9]. However, it is very seldom that *otaku* introduce themselves to others in public because of the negative impression of the term "otaku". Basically it is hard to find real *otaku* and interview them. Especially this research needed to ask interviewees about their feeling of love and sexual desire for two-dimensional characters(see Figure 2 [10]), it was even more difficult to find proper interviewees. In this interview survey, the author approached the interviewees through the own friends and students in the working university.

3.1 Interviewees' attributes

In this paper, the interviewees consist of three men (23 years old engineer, 25 years old student and 27 years old office worker), and three women (21 years old student, 29 years old teacher and 30 years old teacher). All of them have their own favorite digital or two-dimensional characters and use those characters daily on their digital devices such as game consoles with Internet/Wi-Fi function, mobile phones, tablets, and computers. All interviewees lived in Tokyo when the interview was conducted. And also in order to see if nationality affects love feeling and sexual desire, two of them are non-Japanese. And the interview was conducted in off-line face-to-face occasion (4 interviewees) and online video chatting (2 interviewees) on 2014. All of them used the term "love" when they explained on how much they liked their favorite characters.

Table 1: Interviewees' attributes

Gender	Age	Occupation	Nationality	Favorite genre
1.Male	23	Engineer	Japan	Game and anime
2.Female	29	Teacher	United State	Mobile App and anime
3.Female	30	Teacher	United State	Game and anime
4.Male	25	Student	Japan	Game and anime
5.Male	27	Office worker	Japan	Game
6.Female	21	student	Japan	Manga and anime

3.2 *I can discern between what is real and what is not real*

One of my interviewees, a 23 years old man(1), told that he could not imagine his life without his favorite female gaming character. He felt she (his favorite character) made his days happy and joyful. He meets her basically in his computer and game consoles. He works as an engineer in the software company on weekdays. He cannot use his time for meeting her on weekdays, but still he constantly check how she is by his devices every day after work. On weekends, he spends long hours with her, such as talking (typing words) with her, managing her schedule, thinking about her costumes and so on. And when the off-line events are held in real occasions (off-line meetings, special concert events and movie premieres etc), he goes out with her (his devices) and communicate other fans. Another 29 years old female interviewee(2) told that she said good night to her favorite male characters on her iPad just before going to sleep every night. She has some favorite male characters on her mobile apps. Those characters are designed to say what women want to listen from their boyfriends and partners with very popular voice actors' voices. She taps her favorite characters' icons on the screen and listens all of their good-night messages on the bed. She is very busy at working on weekdays and weekends and very tired physically and mentally in the evening. Seeing their handsome faces and listening their cool voices made her relieved and calm, she said. In this interview survey, all interviewees use their favorite characters basically in the similar way. All of them have strong empathy and attachment for their favorite characters. They answered the characters played the role of a boyfriend or girlfriend in their minds. Then, how is their off-line life? Can they see differences between the real and the virtual?

Figure 2: Virtual pop stars *THE IDLEM@STER*

Of course I can discern between what is real and what is not real! All of them emphasized they knew differences between the real and the virtual. Four (1,2,3,5) of them explained that they had their human romantic partners to prove their words. Two(3,5) of them don't say anything to their partner about their favorite characters because of its negative image. Other two(1,2) of them told their partners about their favorite characters and they enjoyed playing their favorite games together. They have the romatic relationship with human partners, and also enjoy the virtual relationship with two-dimensional characters separately from the romantic relationship in their off-line world. The interviewee(1)'s girlfriend has her own favorite male gaming characters and devotes her time and money to play with them. The couple understand each other in terms of playing games and they have never had any conflict because of games and gaming characters.

3.3 Sexual desire and two-dimensional characters

lower compared with other countries' rates in 2013 except Italy, US 1.86, France 1.99, Sweden 1.89, UK 1.83, Italy 1.39, Germany 1.41. [10]

[9]Otaku means ardent fans in specified cultural area in Japan, sometimes the term refers only to ardent fans of animation.

[10]http://idolmaster.jp (in Japanese)

Two of the interviewees (5,6) like porn gaming/manga characters. Although one of them (5) has his human girlfriend and have a sexual relationship with her, he likes playing porn games mainly on his personal computer and also likes seeing how characters get to be his virtual girlfriends and how they react in their virtual sexual activities. He said that he felt great satisfaction when he won a game [11]. He said that playing a porn game is almost same as watching a porn video or reading a porn magazine. He added "but in a game, we can interact with characters and find more satisfaction not only sexually but also mentally than a video and a magazine". According to him, he cannot do anything same to his girlfriend and her reactions are very different from what he expects often in their relationship. For him, gaming characters belongs to him completely. On the other hand, his human girlfriend is an independent existence in his life and he does never treat her as his gaming characters. He is aware how difficult it is to keep the relationship with a human girlfriend than with two-dimensional characters.

The other interviewee (6) told very similar explanation of the reason why she likes porn manga characters. She has never had any boyfriend and any sexual relationship in her life. She said she had a hope to have a boyfriend as well as her friends and enjoy a romantic relationship when she was in her teenage. However, her female friends always complained about their boyfriends and she saw some of her friends had very sad experiences in their relationships with boys. She disappointed at the reality and lost her motivation to have a boyfriend. She said "But I am a human being. I have sexual desire as a normal human being". When she feels sexual desire she reads "boys love" magazines. "Boys love" as one of manga genre draws a romantic relationship between male characters, including sexual depictions. It is generally said that most of fans in that genre are women. She is one of them. She said that those characters did never hurt or bother her and those sexual depictions were very beautiful. She also knows how difficult it is to have a romantic relationship as she wishes, as well as the interviewee (5). Having favorite two-dimensional characters, even including seeing its sex depictions, could be deviation in order to make their lives happier and sometime to escape from the difficult reality.

4. NORMAL OR INSANE?

Although this research shows some people who have special attachment (what they called *love*) for two-dimensional characters through the interview survey, there might be still questions: How is it possible for us to love fictional characters and to get sexual gratification? What is love? What is sex? Or, are they just insane?

4.1 Insanity

Insanity was constructed socially in the modern era. Modern medicine as the political institution and function excluded the man of madness from society to govern it effectively [11]. Until then, madness or insanity did not mean mental disease. Rather, before the modern era, insanity was described as*Eye* available to see even unreasonable things in Plato's *Timaeus*, or as someone who have tragic nature such as Shakespeare's *King Lear* or Cervantes' *Don Quijote* [12] [13] [14]. Suppose insanity is the status of being"non-reason", who can define"reason" and "non-reason"?

Indeed, insanity has always been associated with eroticism derived from libido since the ancient time [15] [16]. However, all of us feel eroticism and that is why history of mankind has been continued. Eroticism is fundamental not only for reproduction but also for relating to one another, being together and competing with each other and developing civilized society [15] [16]. When thinking about sexual activities or eroticism, we come up with the idea of sexual acitivies for reproduction. The idea seems to be very reasonable and rational. Loving two-dimensional characters and feeling sexual desire for them are apparently non-productive and some people have a feeling of dislike to those phenomena. For example, the BBC TV program dealt with strong attachment for two-dimensional characters as one of reasons why Japanese society faced the low fertility.

On the other hand, we already know there are many sexual activities not as reproductive activities [12]. Many young people get to be a big fan of pop stars or movie stars and wish to be a boyfriend/girlfriend of stars. We have already seen many phenomena similar to the case described by this research in our daily lives. Moreover, when two-dimensional characters respond to users in the game, the interaction between users and digital characters gives users satisfaction to some extent, or at least users can feel those characters close to them. Given that love is reciprocal based on the social exchange theory, it seems to be more reasonable and rational than just seeing pop stars on TV monitors without any interaction.

4.2 Eroticism

When saying about feeling a sexual desire for two-dimensional characters, it might sound very strange and some people might feel disgusted. However, we could explain it from the perspective of eroticism. Eroticism is differentiated from sexual activity for reproduction. However, it is still brought to us based on sexual desire for reproduction [16]. In other words, although eroticism doesn't have any reproductive purpose, it strongly connect to profound desire for *life* and our fear at *death*. Eroticism gives us a great pleasure of life and a feeling of awe at death. As Bataille described, all of us are discontinuous beings and long for being continuous being [16]. Death clearly tells us that we cannot exist forever and we are discontinuous beings.

Fictional characters and two-dimensional characters are nothing to do with death [13]. They are basically immortal and emancipated from death. As long as its controlling system is maintained, they can be a continuous existence, even if their users change. They are always beautiful and do never get old. Furthermore, sexual desire and eroticism are driven by desires for violence and violation. Human beings as dis-

[11]In his explanation,"win a game" means having a sex with characters in a game

[12]Masturbation would be one of the typical non-reproductive activities. And as many researches have already revealed, this non-reproductive activity can be observed not only in human beings but also in many animals.

[13]As we see the case of AIBO, when companies managing characters' systems stop or cancel their services, those characters might"die".

continuous beings try to get continuity by violating other's existence [16]. In that sense, fictional characters as a continuous existence allow people to violate its existence limitlessly and easily. Some people could see a hope to become continuous beings through characters as continuous existence. In the BBC video as noted above, the interviewees told the reporter that they could feel like they were teenagers and also they relived their high school days when they were playing with their favorite characters. It might be called nostalgia. However, nostalgia is not unrelated to a wish to life, and to tracing the life continuity inwardly. Having sexual desire for non-productive purpose is a very natural phenomenon for all of us. Technology created new objects for loving or feeling eroticism, and which fit our profound desire for life and fear at death.

5. CONCLUSIONS

This paper shows loving or feeling sexual desire for non-organic objects could be possible for all of us based on the interview and a philosophical perspective of sexual activity. However, still some might say they don?t have ability to love real human beings. Fromm described in his writing *love as an Art* that "requires knowledge and effort" [17]. In modern society, people want to be a person who would be loved by others, and the choice of a "love object" is a main concern in intimate relationship. In the case we love someone, we need to make more effort and to acquire knowledge (art) in order to *love*. Even if we love someone strongly and passionately, we cannot always be loved by him/her as we wish. Sometimes, unrequited love tortures us and hurts our heart. Basically human interaction and relationship are established based on shared expectations and reciprocity between us. It is very difficult for us to be in love with each other under the absence of shared expectations and reciprocity. There is no wonder if someone loves a fictional character as a partner, in order not to get hurt or depressed.One of female interviewees says her digital characters always tell her exactly what she wants to listen, and her characters do never bother and disappoint her. On the other hand, some of interviewees have real human partners in addition to their virtual "partners". According to the interview survey, for the interviewees, having virtual partners separately from real human partners is a way to heighten the quality of their lives. Two-dimensional characters are used not only for disappointment at love but also for enhancing the quality of life including sexual life.

Technology opened up more opportunities to find better partners and objects to love . In addition, technology related to sexual activities currently goes far beyond digital characters. Technology influences our intimate life more and more. Today, many wearable devices to experience virtual sex have come into the market. And also robots designed for having a sex with human beings are being developed rapidly. People might claim having a sex without love or without reproduction is just totally pointless. Or, having a sex with robots is totally "insane". But apparently there are big market needs, and modern technology seems to be able to satisfy them. This paper explored why people want to have technology for satisfying sexual desire from the perspective of insanity and eroticism. Taboos could become the ordinary over time, as we have already seen in our daily lives. But still there are many taboos in our intimate lives, especially in sexual activities. Technology might change taboos to the ordinary and also show the different interpretation of love and sex in the future.

6. ACKNOWLEDGMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006.

7. REFERENCES

[1] S. S. Iyengar and M. R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology*, 79(6):995, 2000.

[2] B. Schwartz. The paradox of choice: Why more is less. Ecco New York, 2004.

[3] Sex and love: The modern matchmakers, *Economist*, 11th Feb 2012. Available online: http://www.economist. com/node/21547217 (Last checked, 5th July 2015)

[4] E. J. Finkel, P. W. Eastwick, B. R. Karney, H. T. Reis, and S. Sprecher. Online dating a critical analysis from the perspective of psychological science. *Psychological Science in the Public Interest*, 13(1):3–66, 2012.

[5] P. S. Dyrenforth, D. A. Kashy, M. B. Donnellan, and R. E. Lucas. Predicting relationship and life satisfaction from personality in nationally representative samples from three countries: the relative importance of actor, partner, and similarity effects. *Journal of personality and social psychology*, 99(4):690, 2010.

[6] W. D. Ross et al. *The Nicomachean Ethics of Aristotle*, volume 546. Library of Alexandria, 1963.

[7] Aristotle (Takada, S. translated to Japanese) *Ethica Nicomachea*. Iwanami Shoten, Japanese translation edition, 2009.

[8] P. M. Blau. *Exchange and power in social life*. Transaction Publishers, 1964.

[9] J. S. Coleman and J. S. Coleman. *Foundations of social theory*. Harvard university press, 1994.

[10] Cabinet Office, Government Of Japan. *White paper on measures on declining birthrate 2015*. 2015. Available online: http://www8.cao.go.jp/shoushi/shoushika/whitepaper/measures/w-2015/27pdf.(Last checked, 5th July 2015)

[11] M. Foucault. *Madness and civilization: A history of insanity in the age of reason*. Vintage, 1988.

[12] F. M. Cornford. *Plato's cosmology: the Timaeus of Plato*. Routledge, 2014.

[13] G. Bullough. *Narrative and Dramatic Sources of Shakespeare: Major tragedies. Hamlet, Othello, King Lear, Macbeth. Volume VII*, volume 7. Columbia University Press, 1973.

[14] M. De Cervantes. *The Adventures of Don Quixote*. Jaico Publishing House, 2015.

[15] G. Bataille and G. Bataille. Les larmes d'éros. 1971.

[16] G. Bataille, G. Bataille, G. Bataille, and G. Bataille. *L'érotisme*. Ed. de Minuit, 1957.

[17] E. Fromm. *The Art of Loving: The Centennial Edition*. Bloomsbury Publishing USA, 2000.

Systematical Follow-up in Social Work Practices

Sheila Zimic
Researcher and method developer
Association of Local Authorities
Västernorrland County, R&D Unit
+46 70-348 26 26
sheila.zimic@kfvn.se

Rolf Dalin
Method developer and statistic support
Association of Local Authorities
Västernorrland County, R&D Unit
+46 70-674 00 94
rolf.dalin@kfvn.se

ABSTRACT
In this paper, we explore the meaning of a specific follow-up system in different social work practices. We want to gain understanding about if the system can be an empowering tool for the practitioners to learn in their professional work. We interviewed four practitioners in different teams in order to find out how they describe the follow-up system as a tool, how they describe the knowledge they produce with help of it and what this means to them.

We have found that the practitioners describe the tool as an easy way to present results which are often described as 'facts'. The teams mainly use their specific follow-ups to present their work to others and in relation to that the tool serves a purpose for the team to gain credibility and be considered as professional. Their use can be understood in light of Evidence Based Practice and performance measurement as a discourse which sets limits and have effects on professional learning.

Categories and Subject Descriptors
K.4.3 [**Organizational Impacts**]: Computer-supported collaborative work: use of follow-up system, learning in professional team

General Terms
Measurement, Documentation, Performance, Design, Reliability, Verification.

Keywords
Systematical follow-up, social work, knowledge, learning, profession, work ethics, performance measurement

1. INTRODUCTION
The Research and Development Unit at the Association of Local Authorities in Västernorrland County, Sweden, have since the year 2010 been supporting various practices in municipality based social services to use systematical follow-up. The specific support for this systematical follow-up model is called LOKE (local evidence), which can be described as a local strategy for knowledge based practices in social services [1].

As the authors of the LOKE-report [1] emphasize, from the mid 1990's there was a severe criticism regarding the practice of social work in Sweden. The main critique had to do with the insufficiency to use scientific knowledge and to use systematical approach in documentation. It was argued that there was a lack of knowledge regarding the effect of interventions – did it or did it not help the client? This enhanced the interest for evidence based practices and for the idea of being able to make ethically optimal decisions for clients', based on 'evidence'.

Later the Swedish *National Board of Health and Welfare* declared that local strategies like municipalities' local follow-up and evaluation activities are important steps in the attempt to make the social work evidence (or knowledge) based [1]. LOKE is one example of such a strategy. It means that the model can serve as a foundation to evaluate if there are any changes in clients' life circumstances after they have participated in one or a combination of activities or interventions.

The aim of this study is to *gain understanding about if our approach for systematical follow-up in different practices in social work can be an empowering tool for the practitioners to learn in their professional work.*

We argue that there is a need to better understand the possibilities and limitations of the systematical follow-up, especially as evaluations and follow-ups are gaining more and more interest in the society [2]. We want to create an understanding of what kind of *knowledge* is strived for in evidence or knowledge based practices, what it means for the practitioners when it comes to *learning* in their profession and how arguments regarding *ethics* in social work are used in this context.

Our attempt is to explore this relationship by studying four different teams which use our approach of systematical follow-up. We want to understand the circumstances and implications with the specific tool each team is using; and we want to gain better understanding of what kind of knowledge they produce with the use of the tool. How the practitioners describe the tool and the knowledge which they produce and what it means to the team.

In order to explore this we pose two research questions:

i) How do the practitioners describe the reifications / visual representations that the tool produces?

ii) What does the use of the tool mean for the team?

1.1 Background
Our approach to follow up on social work results has been bottom-up or grassroots, meaning that analysis of a team's mission and theoretic base as well as the design of the team's data collection form, with our assistance, is done by the team itself. This fits into a management culture that is associated with

continuous organizational learning, and by which "individuals are empowered to learn each and every day" [3]. The purpose of the approach is to create a system that supports the continuous professional learning of the social worker.

We made three main choices that determined the path along which we took off in our development of support for systematical follow-up. The first was to follow the LOKE model mentioned above. Based on the idea in LOKE, we designed a process in three steps, where a crucial first step is describing and analyzing the purpose and rationale of an operation or team in a municipality's social service organization. Based on this analysis or construction, a form is developed for registering data about each client in the team's operation. These data include background, needs/problems, received therapies or interventions, and outcomes for the client.

The second choice was to use a web based survey system called Netigate (www.netigate.net). Reasons for this choice were that it included required data export and report formats, and provided analysis tools which we thought would meet the basic data analysis needs of practitioners.

We soon realized that a team's use of systematical follow-up designed by themselves for their own need, starts a process among them, consisting in understanding possibilities that systematical follow-up could offer. So our third choice was to start and support this process and make the system flexible.

2. THEORETICAL FRAMEWORK

We use a socio-cultural theory of learning [4]. This means we understand learning as a crucial aspect of participating in the world. Learning is a part of social practices in which the learners, the active participants, are producing meaning. According to Wenger [4], focusing on participation has broad implications for what it takes to understand and support learning. It has implications to individuals, communities and organizations.

Engaging in a practice of work means experiencing, participating, and contributing to a practice of a work place community. Billett [5] uses the concept participatory practices and points out that, on one hand, workplaces intentionally regulate individual's participation through activities and interactions that workplaces afford learners, on the other hand, individuals actively choose how to engage in their workplace practices. Both dimensions are important and influence the way individuals participate and learn through work. From this perspective learning and participation in work are inseparable.

2.1 Participation

Participating in a workplace setting means that practitioners make sense of what their work is about, they produce meaning about the specific work together with other participants (colleagues, clients etc.). It also means to become something – to create a profession identity. In this sense, learning and knowledge is institutionalized in social practices. In order to understand learning we must also understand the discourses of learning and their effects on the ways we design for learning [4]-p. 9. From this theoretical perspective, practice as a concept focuses on a process by which we can experience the world and our engagement with it as meaningful [4]-p. 51. Participating in the world as active subjects, in social practices, is about production of meaning, according to Wenger by a process of negotiating meaning, which involves participation and reification that form a duality fundamental to the human experience of meaning and thus to the nature of practice.

The use of the concept "negotiation of meaning" implies that it is a productive process. It also implies that meaning is not constructed out of nowhere. Negotiation of meaning is always related to already construed meaning, in its specific social and historical contexts. Meaning appears fixed by its relational position towards other elements. However a system of relational positions can never be fixed or static because relations do not form closed systems. Meaning is thus never fixed or static even though it might appear as such [6].

Participating in social practices constantly changes the situations to which it gives meaning and affects all participants. From this perspective, there is no sharp distinction between interpreting and acting, doing and thinking, or understanding and responding. All are parts of the ongoing process of negotiating meaning. Participation is thus a process which always generates new circumstances for further negotiations and meanings. It constantly produces new relations with and in the world [4].

2.2 Reification

Wenger [4] describes the other process, reification, as "the process of giving form to our experience by producing objects that congeal this experience into 'thingness' or to "treat (an abstraction) as substantially existing, or as a concrete material object" [4]-p.58. Reification can refer both to a process and to its product. The process and the product always imply each other. The products of reification are not just material artifacts; they should be understood as representations of human practices or symbols of vast expanses of human beings [4].

When something is reified, its meaningfulness is always potentially expanded and potentially lost. "Reification as a constituent of meaning is always incomplete, ongoing, potentially enriching, and potentially misleading. The notion of assigning the status of object to something that really is not an object conveys a sense of mistaken solidity, of projected concreteness. It conveys a sense of useful illusion." [4]-p.62. As Wenger discusses, reification and participation cannot exist without one another. They are interdependent. Reification always rests on participation and participation always organizes itself in reification.

In relation to our study, members of a social services work team embody a long and diverse process of participation. The systematical follow-up of certain practices embodies a long and diverse process of reification. It is in the convergence of these two processes, in the registering of information about clients' as structured data and interpreting the statistical representations, that the negotiation of meaning takes place.

2.3 Appropriation

The practitioners of social work are not designers of all of the rules, documentation and systems they use. In fact, a large portion of reification involved in their work practices comes from outside the communities of practitioners. None the less, the reification needs to be appropriated into a local process in order to become meaningful [4].

To have an idea of how a tool is to be used is not the same thing as how it is really used in practice, in each specific context. The concepts of reification and appropriation suggest that a mediating tool or artifact, for example a follow-up system, can take a life on its own, beyond its context of origin. In our interpretation and use of the LOKE-model, we can see an example of this:

i) we apply the theoretical model outside the practices of its original use,

ii) we have a somewhat different theoretical approach to the understanding of learning and knowledge and how one could understand systematical follow-up in relation to learning and creating knowledge,

iii) we have a different approach to knowledge feed-back, and

iv) we use a different IT-system than what has been used in the original set up where LOKE-model was developed

Thus the concepts of *participation*, *reification* and *appropriation* are useful in our study.

We have a fairly clear picture when it comes to the *aim* with the use of a systematical follow-up in social services, which is to get to know the associations and relationships between clients' problems/needs, the work teams interventions and the outcomes for the clients. However, our knowledge is limited as to what this means to specific teams.

3. NETIGATE AS THE STATISTICAL TOOL FOR FOLLOW-UP

Generally speaking, statistical data analysis results are mathematically functions of data of some specified structure. One such data structure is the data matrix with cells with data arranged in rows and columns. Each row then corresponds to a unit of investigation and each column corresponds to a variable, which is an attribute shared by these units [7]. The attributes can be numerical or categorical, and these variable types restrict and guide the types of analysis that can be used.

Statistical data analysis is basically made by arranging/rearranging the data and computing values of different characteristics of data, known as statistics. Such results are mostly displayed by various visual representations, such as tables or graphs, to be interpreted in analysis reports. The data analysis is typically performed by the use of computers and statistical software packages, of which SPSS (http://www-01.ibm.com/software/analytics/spss/products/statistics/) and Minitab (http://www.minitab.com/en-us/) are two well-known examples. They have both been in constant development since the 1970's. Such software are examples of artifacts which are mediating tools for representing the information stored in data and offering meaning [8] through those representations. Prerequisites for being able to use this kind of tools are training both in statistics as a subject and in the use of the statistics package.

Over the last decades web-based services for carrying out surveys have been developed. These services provide the survey form and the data but also basic tabulations and charts illustrating results. Today's generation of these systems deliver the collected data in different formats, and provide tools both for basic data analysis and for creating reports. The skills needed for gaining useful information from such systems, besides being able to use a limited number of functions in the web-based statistical tool, is the ability to pose relevant questions to the data at hand and to interpret the results in terms of the present context of work practice.

We understand that interpretations of observations within social sciences are context dependent. See e.g. [9] who write "Depending on the circumstances and the situations wherein they are observed, social phenomena have multiple empirical appearances". Therefore, considering the range of types of social work operations we support, and the importance of relevance for those participating in the workplace activities, we decided to organize the follow-up support using a web-based survey tool called Netigate (www.netigate.net), which has the functions mentioned above. Obtaining quick results from statistical data analysis has been made possible in Netigate and similar artifacts, and focus can be directed to interpretation and making sense of results from the analysis and to the possible practical implications in the work-team's community of practice.

Obviously a limited range of analyses are possible in those systems compared to statistics packages. However, being aware of these limitations, we expected team members, using the artifacts built into Netigate, to be able to construct tables and diagrams about their clients' problems, needs, treatments and outcomes relevant to the specific context of each operation's practice. We also expected that using the Netigate system would facilitate the team's communication through oral presentations, written reports and discussions – vertically and horizontally – and thereby support learning in the team, which is a community of learning and practice [4] as we see it.

Put in other words, we believed that a strategy of making interpretation and communication of results an internal process of the social work team, would favor relevance and usefulness of the feedback in the operation. Using traditional statistical packages would be more demanding and not feasible, as we see it, since it would require that data was analyzed and interpreted outside the operational unit's context.

We expect that one important characteristic of this process is that the follow-up system is local in the sense that it is there because of the relevance to their specific operations. This makes the data directly connected with its context, which may guide them in using statistic results, by helping to make interpretations and reflections on tables and diagrams in the light of theories and strategies which they use about their work with clients.

We interpret the statistic tool, Netigate, as a *cultural tool*, according to Säljö's terminology. It means that tools (both physical and intellectual) are temporarily fixed human externalizations of knowledge which provide us with "meaning offers". They suggest activates in which we can create meaning and understanding but it requires active subjects to reconstruct what the meaning could be [10]. In Wenger's terminology the statistic tool can be said to contain several reifications produced in processes such like statistics and computer science [4]. The visual representations, in which the statistics results are represented, such as tables and graphs, are also a type of reifications.

For us it is interesting to find out how the statistical visual representations made available in the specific follow-up system are used as *inscription device* in the reifications of teams' practices. An inscription device could be understood as any instrument or set-up which can display "the end result" of a long process of participation. An inscription device provides a visual display of some sort and helps to mobilize resources of for example a text, into visual representations of a text [11].

Another central thing for cultural tools is that they make it possible to build up External Symbolic Storage Systems (ESS) [12]. With help of ESS people can store information and traces of their experiences outside their bodies. So inscription devices and ESS are according to Säljö [10] important when it comes to the capacity of storing information (which also is of a more indelible character then the human mind), to communicate with other people, but also to think with and to work with.

4. THE MEANING OF A DISCOURSE

According to Wenger *learning* is situated. In order to understand learning we also have to understand the discourses in which learning is constituted [4]. Discourses (or whatever we choose to call a type of normative meaning-making practices) are also a part of information systems design [13], meaning they are situated in social and historical contexts - never value neutral, static or fixed. There is a need to make visible the preconceptions and taken for granted images of the roles an information system is given in a context [14].We argue that this is important to have in mind when we go about exploring the relations between teams' use of their specific follow-up systems and *learning* in their profession.

Using quantifiable measures to observe, to learn and to get an overall picture of a practice/ a phenomenon is not something new regardless of practice of work. However, since 1990's in Sweden a larger focus has been given to structural, systematical analyses in which quantifiable measures are a prerequisite. *Evaluation* has become a trend in public services and in the society in general [2], [15] in which *performance measurement* is a common type.

Performance measurement can be defined as a system for regular documentation, registration and analyses of a work practice regarding specific requirements. The information generated in performance measurements is often quantitative, such as statistics, indicators and "key figures" and can for example be used in presentations, as information to make decisions and internal development of the work practice and it can be used in order to direct and control. The idea with performance measurements is also that it be used by the citizens to gain insight and to be able to make informed choices when selecting a service [2]-p. 11-12.

Even though there are good arguments to the wide range of implications of use of performance measurements or evaluations in general it is also something taken for granted in modern welfare societies and made into a routine to document almost everything. Lindgren compares this activity of constantly providing information for evaluation systems to the analogy of a monster that has to be fed and can never be satisfied. The consequences of use can for example be - large expenses; that it takes a lot of time from the practitioners and that it takes time from other activities. [2]. Thus, there are also arguments to question the 'taken for granted' use of evaluations.

4.1 Evidence Based Practices

One of the fundamental factors, drawing the focus to performance measurements in the social work practices, is the orientation towards Knowledge or *Evidence Based Practices* (EBP) [2]. It can be understood as something that legitimizes and rewards the strategies to use performance measurement in social work. In this sense we talk about discourse as a framework that sets boundaries for what is accepted or not, what is encouraged or not. We see the discourse regarding knowledge or evidence based practices in social work as one such *framework* or *discourse* of learning [4] and we are interested in its effects on the ways we design for learning.

The National Board of Health and Welfare and *Swedish Association of Local Authorities and Regions* have published several reports and guidelines with definitions and methods regarding EBP in social services [16]-[19]. Systematical follow-up is often emphasized as useful and even necessary tool to achieve knowledge / evidence based social services [16]. The argument is that social services should be of use/benefit for the clients and interventions that are pointless or harmful for the clients should be stopped [17]. This relates to one of the foundations in social work practices, namely the *ethics* in social work which are i) that the practitioner should avoid to harm clients; and ii) the work with clients should be performed in relation to scientific knowledge and reliable, experienced knowledge [19].

The use of a follow-up system should, according to Oscarsson [19], provide the practitioners with reliable, experienced knowledge and contribute to EBP-practices. In this case we argue that the ethical argument – *How do the social services know that they are useful and not harming individuals who take part of the services/interventions?*; is maybe the strongest argument legitimizing and rewarding systematical follow-up in social work practices.

In the argument of EBP there is an idea that the practitioners should be able to communicate the result of their work. The performance measurement here is to be able to show that social work practices are of use or benefit for the clients. (How do you know you have done a good work? Show it!) It is supposed that through systematical and structural documentation the social workers can 'really see' what they are doing and therefore be able to evaluate their work. Following the logic of this argument it would mean that if the practitioners do not undertake some systematical methods it can be considered as if they are 'just guessing' which interventions are useful or not.

Systematical follow-up is today a common practice in social services and thus a reality for many of the practices in social work services which R&D Västernorrland works with. Our approach with systematical follow-up relates to the existing EBP-discourse and the arguments regarding ethics mentioned above. In our analyses we try to focus on understanding practitioners' use of the tool (LOKE/Netigate) in relation to *learning* in their professional work and the design (as the effect from the discourse) for their learning. This means we try to see our approach of systematical follow-up in the context of what it is designed for and how it is used by the social work practitioners in a few social work practices.

5. METHOD

We performed semi-structured interviews with four practitioners in different teams in social work. The interviews were conducted during the spring 2015. Two of the interviews we (the authors) participated in together and we conducted one interview each on our own. The interviews lasted about one or one and a half hour each. They were recorded and transcribed. The interviewees are here referred to with fictional names. Also, we chose not to specify which municipality they come from in order to minimize the risk of them being identified. In most municipalities there is only one team providing the services in specific areas of social work.

5.1 Description of the teams

We selected interviewees among those who had several years of experience of using the specific follow-up system, and who were now involved in municipality service operations which had been started as a development project using the LOKE model from the start. The reasons for this were that our interest was in the processes which started by introducing systematical follow-up in these teams.

One interviewee was involved in cases with violence in close relationships in a small municipality, and was the only employee doing the work and using the follow-up system. The operations needed to become more structured and visible for colleagues and

for the public and political level. The follow-up system had been in use for two and a half years. Another informant has been in charge of an open unit for counselling to young people. The team has about six employed counsellors and after a formative evaluation they have used the follow-up system for more than two years. Yet another informant has a key position in municipality employment activities. It is a start-up project subject to overall evaluation. Its structure is complex since it is a co-operation between governmental and municipal authorities. The fourth informant is in charge of a central unit in one municipality with a special task to promote attendance in primary and secondary schools, mainly by addressing individual cases.

5.2 Analytical strategies

We analyze the empirical material with help of the theoretical, analytical concepts presented earlier in this paper (see sections 2-4) with guidance of our research questions. We try to find descriptions of the *tool* and the *profession* in interviewees' articulations.

More specifically, we are inspired by discourse theory [6], [20]. Even though we are not attempting to do a discourse analysis we find it useful to try to find elements which could be related to one another and make chains of meaning – *equivalence chains* regarding the tool and what it means to the practitioners.

6. ANALYSIS

We could see three main equivalence chains – *To corroborate an argument*; *To make work visible* and *To achieve goals*. These are not separated from each other. We rather interpret them as chains of important elements in teams' participatory practices, related to their context which is situated in a discourse. We present the first two chains of meaning: *To corroborate an argument* and *To make work visible* in the analysis. The last - *To achieve goals* is presented in conclusion since we consider it to be summarizing our findings.

6.1 To corroborate an argument

To *easily*, *perspicuously* and *quickly* be able to present results seems to be an important strategy for the teams. All interviewees describe the tool in terms of being able to "easily show", to present their team's work in figures, bars and tables. These are the visual representations provided by statistical representations, which are described with words such as: *raw facts*; *hard facts*; *hard statistics*; *clear*; *visible*; *real/reality*; exact (*not fuzzy*) etc.

All of these words imply that the meaning given to statistical representations is that they objectively 'reflect a reality'. This given meaning also relates to certain status regarding what we see as knowledge and truth. Our interpretation of the interviewees expressions such like:

It's not just us driveling; we know what we talk about (Alice, School Team)

and

Show that we're not just making up (Maria, Youth Service Team)

is that the statistical representations are given a high status in relation to the meaning of knowledge and truth. According to Latour [11] the construction of facts is a collective process, which among other things involves the use of already established facts and artifacts – black boxes. The black boxes can be, but are seldom, reopened and questioned. Statistics and statistical representations can be described with Latour's terminology in the sense that statistics and statistical representations are established, taken for granted facts and artifacts in our society. Säljö [10] would describe it as *cultural tools* - temporarily fixed human externalizations of knowledge which provide us with "meaning offers" and suggest activates in which we can create meaning. But it requires active subjects to reconstruct what the meaning could be.

One of the interviewees, Alice, expressed the following in relation to the use of statistical representations:

An image's ability to get through to a person should not be underestimated. To read a text - seventeen, fifteen per cent, but to get an image. If it is bars or circle diagram. (Alice, School Team)

Hence, there seems to be something very fundamental, something taken for granted by the use of the statistical representations in order to establish knowledge and facts, something that goes beyond a specific discourse. So, what kind of knowledge are the social work practitioners aiming to produce?

As we could see the interviewees clearly expressed that the teams used their follow-up systems to corroborate their arguments or establish 'facts', not in a scientific meaning but rather establish the facts of their reality. The interviewees stress that the usability with the follow-up system is to be able *to show results*, *to make visible the work they are doing*.

To themselves, it is valuable to make visible both what they already experience, what they meet in their everyday practices with clients but also to see unexpected things when it comes to characteristics of the clients. So the knowledge they produce is mainly to construct a *representation* of the team that the practitioners themselves recognize and are acknowledged in.

The process to establish a 'generalized representation' of the teams work includes participation in decisions regarding what is considered to be important to measure and to make visible. How usable the tool is for the team depends heavily on how it is *appropriated* [4] by the team. It depends on how they interpret and make sense of its value for the teams work. It depends on what they decide is important to make visible and how they choose to represent their work in terms of structured information. One of the interviewees says:

Then you can make people do it, if they can see the benefit in the end. That we actually get results we can present about our team and that the projects advances.

…. Because, that's what changes you. When you can see the results, what we get out of it. (Diana, Employment Support Team)

As we interpret it, the usability is achieved when the team gets to design the system. In the participation of design they can make sense of the value of the follow-up system and they can choose what to represent and what not to represent. In this way they externalize/reify [4] their practice of work in structured information in a questionnaire which is the instrument in their follow-up system. The choices of structuration are visible in following statement:

We are not an operation dealing with emergency situations. We are preventive, then the interventions should reasonably be quite delimited, in order to get a new flow. But we do have those who are here a lot. And for a lot of things. And it's a long process. And then we have decided – if we have not had contact for a month, then we have to make a new registration. Otherwise we will get incredibly misleading statistics. Because it is madly important to show figures, for politicians and managers. A figure like that would dig our ditch. (Maria, Youth Service Team)

In this statement we can see how the team choses to adapt the tool in order to represent the teams operation as a *preventive* operation. They are categorized as preventive operation team and they express that they have to show that in figures.

The choice of the Youth Service Team's structuration of how the information should be registered is not just to register as many cases as possible. They articulated other strategies which could have been used if 'quantity' was the only aim. The team also wanted to show that they work as a team and co-operate in the sense that several of the practitioners can meet the same client to help him/her with different things. In the end it is their choices, their negotiations and their decisions that constitute, set frameworks for, the 'generalized representation' in relation to what can be gained from the follow-up system.

The system gives them the possibility to shape their own representation with strategies such as: *it is what we decide*; *do not have to show everything*; *pick out the interesting parts*. Another important aspect of the negotiation and renegotiation of the 'generalized representation' is the *flexibility* of the system. All of the interviewees expressed that it is an advantage that the system is not fixed. If they discover something, either in their everyday work with the clients or in the information generated by the follow-up system, they can make changes in the system.

Finally, this also indicates the point that their structured information in form of statistical representations cannot really be interpreted as if they are 'objectively reflecting the teams reality', since the team itself choses what to show. Regardless of who makes the statistical representations – they are never 'objective' because it is always someone who choses what is important to represent [2]. When the team presents their statistical representations, the representations are interpreted as 'facts' or 'evidence' which sets the rules in relation to what legitimation the team is ascribed by others. We would argue that the teams in our study appropriate their systematical follow-up in a very strategic way, to gain legitimation, to gain resources and status such as to be considered as professional.

6.2 To make work visible

Some of the outcomes of the teams' use of follow-up system, as we already mentioned, is to create the 'generalized representation' and show their work in a way it will gain credibility. The expressions: *Be acknowledged*; *To show what we've been up to*; *Be visible*; *We can easily see*; *We knew what we wanted to register*; *What we want to show*; *Makes visible what we do*; *Remind of what we do*; *This is what we encounter every day, we can show*; *We work like this* and *Clarify* all seem to be referring to the urge to legitimize their work. In that sense the tool can be strategic and empowering for the practitioners in order to gain status such as: *we are now permanent* or that they are perceived as *professional*.

One of the interviewees says:

I was not in the focus anymore and did not have to stand and answer, keep figures in my head. It circulated around this (points at the printout of statistic results).

Not that I have any problem to speak, but this felt professional. (Tina, Resource Team against Violence in close relationships)

This statement refers to the aspect of External Symbolic Storage System (ESS), which has the function, the possibility to store a lot of information but also to communicate it visually in a condensed, way [10]. What is given legitimacy in the use of ESS could be that it is possible to get compilations of all the cases a team works with. The tool makes it possible for the practitioner to see all the cases at the same time and make informed decisions, in contrast to what is supposed to be the opposite – to base her decisions on a few cases she can remember. We interpret this that 'objectivity' is assumed to be inscribed in the instrument [11]. *Objectivity* is ascribed high status and legitimacy in the discourse of EBP, in which the argument about ethics in social work is of central meaning. Since the work with clients should depart from scientific knowledge and reliable, experienced knowledge – systematical, *objective* knowledge, elicited directly from ESS, are considered to be more reliable and 'more true'.

There are such examples of expressions regarding the follow-up system helping to disrupt some common misconceptions about the clients. One interviewee, Tina, who is working in a support team for women who have been abused or experienced violence, explains how the politicians, the decision makers reacted when she presented statistic results of her work:

They were surprised that so few needed interpreters, how many of them had children. They reacted on that. Almost all of the women, ninety per cent at least, had children under eighteen years. And how will it go for the children? So that was sort of, perspicuous. (Tina, Resource Team against Violence in close relationships)

Sometimes it was the practitioners themselves that gained new insights regarding the proportions of clients with specific problems:

We thought it would be a great many of these children that have a diagnosis. And many do have, but not as many as we thought. And that was clear to us when we registered and received the statistics. It was actually half of them. (Alice, School Team)

When it comes to representing results and showing that the team meets the *needs* of the clients, it seems to be trickier. We can see that there are expressions that the follow-up system only provides certain type of information, certain types of representations of the clients. Even in our approach when the practitioners design their own follow-up system. One of the interviewees expresses that the follow-up system misses out individuals' process including the stories and lived experiences of individuals.

Yes, we have (outcome rates) but it is more performance measures that we have. That we will get this many participants, we shall have forty per cent of the participants should end up in employment or studies. That's a goal we have. And that we can measure in what one has ended up in.

But I think it is interesting to know for the individual. Because, even if you did not get a job or started studying, it can still mean a tremendous difference for the individual. What one has experienced and that one has undergone a very big development as well. I think that is also important to show. So that it won't just be that one is a part of a percentage in the report of results. (Diana, Employment Support Team)

Presenting results of the teams work, it is also obvious that it hints the aspect of ethics – "What are the clients' needs? Who defines the needs?" but also simply how the clients are represented. One of the interviewees uses the word *alive*:

We try to inform. We encounter this right now and we tell about specific examples. One young person comes and tells his/her story so that it becomes very alive. (Maria, Youth Service Team)

And maybe this is what the practitioner at the Employment Support Team means when she says that it is important that their clients do not become just a percentage in the report of results. We interpret this expression that it is important to show that the

clients are not just abstract clients, but real persons. According to the same interviewee both representations are needed – the generalized, structured representation and the specific representation.

Well, it is a combination, you could say. They (the decision makers) *want to see what is earned in money value. That's important. And they want to see that results are achieved. That the performance goals are ensured and that, like here at X that we can get participants to start working or studying. That is sort of a goal. Another goal is that it is a journey when it comes to how close to employment one comes. That one becomes more employable.* (Diana, Employment Support Team)

In a team's appropriation of the tool, the realization of what kind of representations the tool can produce, what possibilities and limitations the tool has is in the end what makes it useful for the team. Realizing the limitations saves the team both time and frustration. It helps to find a manageable way of registering the sort of information the team finds important for them to visualize.

Diana: *And then they want to see the process regarding specific cases. That a person has gone from having very big problems to hopefully being more employable and close to get an employment when they finish here. So we have to find a way to elicit that. But that's the limitation* (with the system). *Sometimes there are wishes to get out more things than what is possible.*

Sheila: *Mm, precisely. So you have to be clear about what sort of*

Diana: *Mm, it's not possible to demand of the case workers to log in all the time and register as soon as something happens in a case. Instead we have realized that an input and an output is what's manageable for us. And that's what's interesting in relation to the goals of the project.*

Then, if we want to follow individual cases we have to find a person's specific journey. We have to find other ways to describe that and we have done that by interviewing persons and they have described their journeys'. How it's been for them. (Diana, Employment Support Team)

The interviewees articulate that there are other ways to show results and this is also a way of emphasizing the complexities that are not easily captured in structured information, such as the individual process or journey.

7. CONCLUSION - TO ACHIEVE GOALS

It is of importance that all four of the teams started out either as projects that had to be evaluated in order to see if they were going to be permanent in the organization. Or in one case it was an area of social work that was not given much attention or resources in the organization and the practitioner had to show to others in the organization what she was working with. This affects the reasons and the way the follow-up is used.

One of the interviewees says:

If you want to make a team to be permanent, then it's good to be able to report what you have been up to. We have done that and we are now permanent (Alice, School Team)

In the discourse of EBP the argument is to measure performance in order to make sure that the interventions in a social work practice are of benefit to the clients [19]. But for whom is this information interesting? The argument takes a standpoint that this type of information should be valuable for the practitioners in order to perform as good work as possible. And it should be valuable and ethical in relation to inform the clients about the effective interventions they could choose from.

However, none of the interviewees talked explicitly about *ethics* in meaning 'measuring in order to improve the work'. And *learning* was not a common word used by the interviewees to describe their engagement with the tool. More commonly used words, to describe the engagement with the tool are - *to show*; *what is required* and *see the benefits*. This does not mean that they did not learn anything. Rather, the lack of explicit verbalizations regarding *learning* says something about how the practitioners perceive the tool.

Our interpretation is that the initial focus for the teams was not *learning* in the community of practice, but rather very strategic for the purpose to fulfill the *requirements* from the management. The focus was on showing (and convincing) the management and decision makers that the team is doing a good work. This sets the frame for the questions that the teams posed to the data at hand. Initially they were more focused on the knowledge requirement from the leadership, rather than on the knowledge the team itself wanted or needed to create.

Since the discourse and design for a team's learning primarily focuses on achieving goals, the *learning* is delimited specifically to that purpose. As we see it there is an imbalance between what is argued in the discourse of EBP and the reality of practitioners' professional work. It seems to be inevitable for the professionals not to bother about being evaluated and rated in relation to their work performance.

Another critical point is that the goals are sometimes difficult to find ways to measure. Performance goals "this many per cent should achieve that" seem to be easier then measuring more abstract goals such as "employability" or how much benefit the clients had by taking part of a team's operation (even so if they ask their clients to estimate and grade the benefit). Sometimes the goals are not so specific or not established which makes it impossible to show if the team had achieved its goals. And sometimes the goals / results were too robust and did not really represent the complex situation of the clients in an adequate way.

But still, one of the interviewees, Diana, says that it is important to measure goals that are set by the leadership since the decision makers are responsible to deliver good welfare services to the citizens. She explains that she has past experiences where the practitioners organized their work in relation to what suited the practitioners best – for example opening hours.

Politicians are elected representatives of the citizens and they have their work ethics in relation to meeting citizens' needs. In best case the top-down goals are well investigated/studied and established in relation to the work a specific team performs. In worst case the top-down goals are not investigated and not so well established, which makes it difficult for practitioners in a team to make sense of them [4].

So the question is – how do we find a balance and make the follow-up system primarily as a tool for learning in a professional team?

The interviewees emphasized some of the characteristics such as the possibility to *build up* the system on their own, in relation to what they decide is important to display, and the *flexibility* of the system - that they can change variables and categories when they renegotiate their participatory practices. In order to make the tool be used for *professional learning* and a team's development of work, it needs to be appropriated fully by the practitioners in the team. It means that the entire team needs to participate in the design of the follow-up system (which they most often already do), they need to find strategies to negotiate and make sense of the

'top-down' goals in relation to their practice, they need to find strategies to make the registration work pragmatically in their everyday routines and they need to be driven by their own knowledge needs and curiosity and pose questions to the data at hand.

However, this is a process which takes time. One of the interviewees, expresses that now they have started to realize that the tool is primarily their own, for the team itself to use as a tool in their work:

And it's not a value for somebody else in first place, but for us in the team. So that we can look closer in to and reflect. Go further. Develop. (Maria, Youth Service Team)

8. ACKNOWLEDGMENTS

Our thanks to the interviewees who took their time to share with us their experiences of using the follow-up system.

9. REFERENCES

[1] Hjelte, J., Brännström, J. and Engström, C. 2010 Final Report*: Lokal Evidens (Loke)! En modell för lokal uppföljning av kommunal öppen och heldygnsvård som riktar sig till personer med missbruks- och beroendeproblematik. Ett SKL-uppdrag.* http://www.umea.se/download/18.21003518141dc801bd7392/1382427676649/LOKE-rapport.pdf (Retrived:2015-07-01)

[2] Lindgren, L. 2014 *Nya Utvärderingsmonstret. Om kvalitetsmätning I den offentliga sektorn.* Studentlitteratur, Lund, SWE.

[3] Clow, J. 2012 *The work revolution. Freedom and excellence for all.* Wiley.Tavel, New York, NY.

[4] Wenger, E. 1998 *Communities of Practice. Learning, Meaning, and Identity.* Cambridge University Press, New York, NY.

[5] Billett, S. 2004 Workplace participatory practices: Conceptualising workplaces as learning environments. In *Journal of Workplace Learning*, Vol. 16 Iss: 6, pp.312 – 324. DOI= http://dx.doi.org/10.1108/13665620410550295.

[6] Laclau, E. and Mouffe, C. 2001 *Hegemony and socialist strategy: towards a radical democratic politics* (2. ed.) Verso, London, UK.

[7] Hervé, A. 2004 In The SAGE encyclopedia of SOCIAL Science research Methods, SAGE, California.

[8] Englund, T. (Ed.) 2004 *Skillnad och konsekvens. Mötet lärare-studerande och undervisning som meningserbjudande.* Studentlitteratur, Lund, SWE.

[9] Faber, J. and Scheper, W. J. 2003 Social scientific explanations? On quine's legacy and Contextual fallacies. In *Quality and Quantity* Vol. 37, pp. 135-150.

[10] Säljö, R. 2013 *Lärande och kulturella redskap. Om lärprocesser och det kollektiva minnet.* Studentlitteratur, Lund. SWE.

[11] Latour, B. *Science in Action. How to follow scientists and engineers through society.* Harvard University Press, Cambridge, MA.

[12] Donald, M. 1991 *Origins of the modern mind. Three stages in the evolution of culture and cognition.* Harvard University Press, Cambridge, MA.

[13] Orlikowski, W.J. and Iacono, C.S. 2001Research Commentary: Desperately Seeking the "IT" in IT Research – A Call to Theorizing the IT Artifact. *In Information Systems Research*, Vol. 12, Iss. 2, pp. 121-134.

[14] Cecez-Kecmanovic, D. 2011 Doing critical information systems research – arguments for a critical research methodology. In *European Journal of Information Systems*, Vol. 20, Iss. 4, pp. 440-455.

[15] Vedung, E. 2003 *Utvärderingsböljans former och drivkrafter.* STAKES, FinSoc Working Papers 1/2003.

[16] Socialstyrelsen (Swedish National Board of Health and Welfare) *2014 Systematisk uppföljning. Beskrivning och exempel*. Article nr. 2014-6-25 ISBN-978-91-7555-194-4.

[17] Socialstyrelsen (Swedish National Board of Health and Welfare) 2015 *Socialtjänsten. Förutsättningar för kunskapsstyrning.* Article nr. 2015-43. ISBN- 978-91-7555-277-4.

[18] Socialstyrelsen (Swedish National Board of Health and Welfare) 2012 *Om evidensbaserad praktik*. Article nr. 2012-12-20.

[19] Oscarsson, L.SKL (Association of Local Authorities and Regions) 2009 *Evidensbaserad praktik inom Socialtjänsten. En introduction för praktiker, chefer, politiker och studenter*. SKL Kommentus, Stockholm, SWE.

[20] Howarth, D.R., Norval, A.J. and Stavrakakis, Y. 2000 *Discourse theory and political analysis: identities, hegemonies and social change.* Manchester University Press, Manchester, UK.

Privacy concerns arising from internet service personalization filters

Ansgar Koene, Elvira Perez, Christopher J. Carter, Ramona Statache, Svenja Adolphs, Claire O'Malley, Tom Rodden, and Derek McAuley
HORIZON Digital Economy Research, University of Nottingham, United Kingdom
University of Nottingham Innovation Park, Triumph Road
Nottingham, NG7 2TU, UK
+44 0115 8232551
{ansgar.koene, elvira.perez, christopher.carter, ramona.statache, svenja.adolphs, claire.omalley, tom.rodden, derek.mcauley}@nottingham.ac.uk

ABSTRACT
Personal service customization, or personalization, is one of the core tools that are being used by on-line providers of information services such as search engines, social media, news sites and product recommender systems to optimize the individual user experience in hopes of attracting and keeping users. In this paper we will examine the user profile models that are used to achieve this information personalization. From a citizen centric perspective, our concerns focus on the degree of privacy intrusion that is implicitly required to determine the parameter settings of the information filter profile and the ethical implications of the personal behavior predicting properties of the user model itself.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues: Ethics, Privacy, Use/abuse of power.

General Terms
Algorithms, Measurement, Performance, Design, Economics, Reliability, Security, Human Factors, Theory, Legal Aspects

Keywords
Personalization; Behavior profiles; Information filtering; position paper.

1. INTRODUCTION
The massive growth of digital data creation, with more than 90% created in the last couple of years, a 400% data collection increase year-over-year in 2012 [1] and almost 1 billion active and indexed websites [2] has made sifting through and ranking of information into the primary challenge for many internet uses. The basic concept behind personalization of on-line information services is to shield users from the risk of information overload, by pre-filtering search results based on a model of the user's preferences.

As such, the motivation behind these systems is ethically sound. The user profile model that is used to predict a user's preferences, however, and the methods by which the data is acquired for tuning it, do raise concerns.

The user profile model, is often derived from past online behavior of the user [3], which is logged with the user account. This data is primarily derived from previous visits to the service providing site, but in some cases may also involve the use of 'tracking cookies' to gather information about the user's behavior on other websites in order to further fine tune the user profile [4]. Other frequently used sources of data for tuning the user profile models include data concerning the behavior and preferences of people within the social network of the user [5]. Leaving aside the obvious ethical concerns relating to the use of 'tracking cookies', tracking of user activity on the service site itself can also produce highly detailed personality profiles, especially when the service provider is a search engine or social media site that is heavily accessed by the user and provides a wide diversity of services. In essence, the process of creating a user profile for the service personalization involves exactly the kinds of privacy invasive data mining that we have previously argued to require strictly maintained informed consent procedures to maintain proper research ethics when employing such data mining for academic research [6,7]. It is therefore ethically highly problematic that the need to maintain an advantage over competing services frequently results in service providers choosing not to inform their users about the personalization methods that are being used. Despite these ethical issues concerning the data that is used for creating the user profile models, the main concern we would like to draw attention to in this paper is not the 'raw data' but rather the user profile itself.

The user profile model is in essence an operationalization of the data mining efforts, built to anticipate the user's behavior, interests and desires. A perfect user model would ideally, from the service provider's perspective, enable the service provider to perfectly predict the decision a user would make for any given choice. If successful, this would in effect produce a Pandora's box of potential privacy violations, just waiting to happen. To find a user's weaknesses, for instance, it would suffice to query the user's profile model with a range of choices and observe the predicted responses. Such an idealized perfect user profile model is of course not (yet) possible, and would require access to data that is not (yet) in the on-line domain. Increased prevalence of internet connected sensors, i.e. Internet of Things, however may change this in the near future.

In section 2 we provide a brief review of information personalization systems and the role of user profile models in these. Section 3 describes the process of data collection for generating person profiles. Section 4 conceptually summarizes the frequently used method of constructing the user profile model from the collected data. Section 5 discusses some of main uses and possible abuses for which the personalization profiles could be used.

2. BRIEF REVIEW OF PERSONALIZATION SYSTEMS

Ranking and/or filtering of Internet search results and Social Media-/News-feeds for increased user satisfaction is in essence the same challenge as that is posed to recommender systems used by the likes of Amazon.com, YouTube, Netflix, TripAdvisor, etc. to suggest items the user might be interested in. Recommender systems emerged as an independent research area in the mid-1990s. These first recommender systems [8] applied collaborative-filtering which matches users who have in the past made similar choices (i.e. given similar ratings, or 'clicked' on similar items) on the assumption that they have similar preferences and will therefore be interested in recommendations for items that these users rated highly. Modern recommender systems use (combinations of) various types of knowledge and data about users and previous transactions stored in customized databases. The knowledge and data about the users is collected through explicit ratings by the users for products (e.g. purchase feedback on Amazon), inferred by interpreting online actions of users (e.g. navigating to a particular product), through monitoring of social networks and social media activity (e.g. Facebook Social Graph) and increasingly through data from personal networked devices (e.g. Mobile phone location data).

The three main classes of recommender systems are:

1. Content-based, where the system recommends items based on their similarity to items the user expressed interest in, e.g. purchased, clicked on, searched for etc., in the past. The similarity of items is calculated based on the features associated with the compared items.
2. Collaborative-filtering, users are given recommendations for items that other users with similar tastes liked in the past. The similarity in taste of two users is calculated based on the similarity in the rating histories of the users.
3. Community-based, where the system recommends items based on the preferences of the user's friends. This is similar to collaborative filtering except that the selection of peers that are used for selecting the recommendations is based on an explicit 'friendship' link instead of being deduced from patterns of similar past behavior. Such 'social recommender' systems are poplar in social-network sites [9].

In practice many recommender systems are hybrid systems that try to balance the advantages and disadvantages of each class [10]. Collaborative and community based systems, for instance, suffer from an inability to recommend items that have not yet been rated by any of the potential peers of the user. This limitation however does not affect content-based system as long as the new item is supplied with a description of its features, allowing it to be compared to other items that the user has interacted with in the past.

A comprehensive introduction to recommender systems is provided in [11].

3. USER PROFILES INFORMATION GATHERING

We will now give an overview of common user profile data collection methods, including discussions regarding the impact on privacy, the growing role of social networks and issues related to trading of data with third-parties. Most of the examples in this section will refer to Google, simply because of its dominant position in information services. Reference to Google's practices is not meant to imply that their practices are any more or less ethically acceptable than any other service.

3.1 Data collection

Data collection about users typically uses a range of different channels. At the most basic level the service provider, e.g. Google, records the immediate interaction of the user with its service, e.g. the search and browsing activity. With respect to this type of data collection, the Terms of Service [12] and accompanying Privacy Policy [13] which Google presents when a new account is created state that:

"When you use our services or view content provided by Google, we automatically collect and store certain **information in server logs**. This includes:

- details of how you used our service, such as your search queries.
- telephony log information, such as your phone number, calling-party number, forwarding numbers, time and date of calls, duration of calls, SMS routing information and types of calls.
- Internet protocol address.
- device event information, such as crashes, system activity, hardware settings, browser type, browser language, the date and time of your request and referral URL.
- cookies that may uniquely identify your browser or your Google Account."

For the most part the information that is collected through the server logs is unsurprising. Probably least obvious amongst this list are the collection of the phone related information, the system activity and hardware settings. It should be noted however that none of this information actually requires that the user has an account with the service provider (Google). Based on the IP address, phone information or other hardware information, logs of search queries that are performed while the user is not logged in to an account could in principle still be linked to the profile associated with the user's account.

For the construction of a behavior profile, tracking of search queries (or more generally the way in which the primary service function is used) remains a core defining element since this is what the personalization must aim to improve to satisfy the user.

Other obvious data that is collected includes the information which users are asked to provide when they sign up to an account. This typically includes: a name, email address, telephone number and possibly even a credit card. Increasingly, thanks to improved face recognition algorithms, users are also strongly suggested to include a photo. Providing of fake inputs for this personal information is often the first action people take when they become

more privacy sensitive. In itself this information is not particularly useful for the creation of a user behavior profile, but it does provide important linking information for associating user data that is gathered from different, nominally independent services.

More interesting and less obvious data which is mentioned in Google's 'Information that we collect' section in the Privacy Policy includes:

"We collect information [when you] visit a website that uses our advertising services or view and interact with our ads and content. This information includes:

- **Device information,** such as the hardware model, operating system version, unique device identifiers, and mobile network information including phone number.
- **Log information,** [as described earlier under 'server logs'].
- **Location information,** [determined] using various technologies, including IP address, GPS and other sensors that may, for example, provide Google with information on nearby devices, Wi-Fi access points and mobile towers."

Often it is not clear to the user which service is providing the ads on a website, nor does the user know what ads to expect on a website before visiting it. The user therefore has no means of controlling which ad-providing service will know about their visit to a particular site. The only way for the user to regain agency and control over consent is to install ad-block software and/or disable cookies, both of which might disable some browser functionality the user may have been interested in.

The methods that are used for collecting data about web-browsing behavior rely on "various technologies to collect and store information [which] may include using cookies or similar technologies [e.g. pixel tags/Web beacons] to identify your browser or device when it visits a webpage." … "We also combine this data among our services and across your devices for these purposes, for example, using information from your use of Search and your Gmail to show you personalized ads."

From a user perspective unfortunately these 'various technologies' appear to all be beyond the control of the user and are mostly hidden so that the user frequently does not know that such data collection is taking place. This makes it very difficult for users to manage the level of information they wish to expose about themselves.

3.2 The role of Social Networks

Social Networks, like Facebook and Google+ play an increasingly important role in user profiling due to the richness of personal data they contain. In many ways a user's Facebook or Google+ page is nothing else than an elaborate exercise in self-profiling contained in a tightly templated structure that facilitates automated data extraction. To further enhance the depth of the user profile information on Social Network Sites (SNSs), users are repeatedly prompted to fill in more background details (e.g. "what was your role when you worked at X), 'tag' more photos and tell their 'friends' about the latest things they are interested in, while the profiling engine listens to their communications. Most important however is the 'friends' network, i.e. the 'Social Graph', itself which directly establishes the network of peers to use for Community-based recommending systems.

In the context of privacy/consent related issues, one of the main concerns with Social Network Sites is the loss of personal control over the information that is provided to the system, due to the bi-directional nature of the network. This was most prominently discussed in relation to image tagging [14] where users can tag other people, revealing their presence at an event without the explicit consent of that person. The same holds true, however, for many other activities on social networks, including the sending of 'friend' requests. Even if the request is declined, it reveals something about both sides of the interaction. This is especially true since it is notoriously difficult to truly delete something from social network sites, where 'removing' usually only means hiding it from other normal Social Network Site users [15]. Further more, it is not at all clear if/how the parameters on the user profile model are updated when data is 'removed' from the social network.

3.3 Trade in personal databases

Since trading of personally identifiable data to third-parties, without the explicit consent of the individual to whom the data refers, is generally considered to be a too severe privacy violation that would have repercussions for the parties doing the trade, such data is commonly not traded. Instead the policy regarding 'Information we share' [13] states that:

"We do not share personal information with companies, organisations and individuals outside of Google unless one of the following circumstances applies:

With your consent

We will share personal information with companies, organisations or individuals outside Google when we have your consent to do so. We require opt-in consent for the sharing of any sensitive personal information. [Such an opt-in may however be included in the Terms and Conditions that users commonly click-sign without reading when they install new apps.]

For external processing

We provide personal information to our affiliates or other trusted businesses or persons to process it for us, based on our instructions and in compliance with our Privacy Policy and any other appropriate confidentiality and security measures.

For legal reasons

We will share personal information with companies, organisations or individuals outside Google if we have a belief in good faith that access, use, preservation or disclosure of the information is reasonably necessary [for law enforcement]."

However in the second to last paragraph they also state that:

"We may share aggregated, non-personally identifiable information publicly and with our partners – like publishers, advertisers or connected sites. For example, we may share information publicly to show trends about the general use of our services."

Since the data that is shared with partner organizations is aggregated and non-personally identifiable (we will assume that this is indeed the case, unlike [16]) it can not contribute very specific data points to the user profiles. It does still hold a lot of value for the tuning of user profiles, however, since data of the type: 'N percent of people with characteristics J and K chose option A'; does help to shape the predicted behaviour probability distributions for 'people with characteristics J and K'.

4. COMBINING DATA INTO PROFILE MODELS

User profiles are most frequently represented by mapping the data in a high dimensional space [17, 18], with vectors denoting the past preferences the user expressed in their observed online behavior. In order to better capture the context dependent nature of human preference, especially in social settings, some personalization systems use context-aware generative models to adjust the multi-dimensional mapping according to context [19, 20, 21, 22]. Based on this multi-dimensional vector representation of the personal data profile, recommendations can then be generated by projecting the set of potential results into the same space and selecting those items that have the shortest distance from the personal data vectors.

When constructing the personal profile model there are a number of choices that needs to be made, foremost among which is the question of how to define the dimensions. What kind of online items, behavior and communications should be classified as being aligned along a single dimension? What should the unit scales be on each dimension? e.g. is the difference between red and green colors more significant than a doubling in size of an object? In some cases the task might literally consist of comparing apples with oranges, the answer to which is obviously context dependent. The quest to solve these dilemmas is one of the reasons why tech companies like Google and Facebook are investing heavily in 'strong AI' research.

Aside from the ethical issues related to the acquisition of input data for the creation of the model, which we discussed in section 3, the user profile model itself also raises some interesting ethical issues. The purpose of the model is to predict a person's preferences, which is done by a process of nearest-neighbor matching in the mapped multi-dimensional space. Any additional information that is inferred from the accumulated input data therefore only exists implicitly as long as no specific search is done. Does the implicit nature of the information automatically shield the model from any claim of privacy invasion, no matter how personal or intimate the inferred knowledge about a person is?

5. USES AND POSSIBLE ABUSES

The primary uses and purpose of user profile models are to facilitate personalization of the information service (Search, News-feed, product recommendation, etc.) to improve the user experience, as well as facilitating targeted advertising to improve click-through and sales rates.

Since the user profile model is in essence an attempt at profiling and anticipating a user's preferences and behavior, one could easily imagine using/abusing the model for any situation that involves personality profiling. If the user profile models were sufficiently reliable, recruitment agencies could simply arrange submit targeted questions to the profile models to identify the most suitable candidates for jobs possible making job interview redundant. Law enforcement agencies might use the user profile models to narrow the field of suspects or use the profile model to predict the actions of a specific suspect. Teachers might submit queries to the profile models of pupils to help them find the most engaging way to present their course material. Viewed from a techno-utopian perspective, the list of beneficial uses appears endless. Viewed from the citizen-user perspective who's personal profile is being analyzed, however, each of these use cases is ethically highly contentious and would require a lot of safeguards to protect citizens from abuse. None of the examples we listed are currently feasible, due to the low fidelity of the model predictions at this time. The use of the user profiles for targeted advertising, however, has already revealed some of the potential pitfalls as shown by the case in 2012 when the Target used this type of data mining to identify and inadvertently reveal a girl's pregnancy to her father [23].

6. INTERNET OF THINGS

One of the reasons why the user profile models still have only limited ability to anticipate user preferences is that the data they are build on is mostly confined to the behaviors people exhibit online. In order to get a more complete profile of a person it will be vital to incorporate data from real-world behavior. The first move in that direction was obviously location tracking in smart phones which could for instance help to disambiguate location dependent context effects on user preferences. Fitness monitors and health trackers (e.g. Apple Health-Kit) are now set to add information about the physiological state of the user.

The main ethical concern that is raised by the introduction of Internet of Things devices as additional data source for the user profiles is the inherent privacy invasiveness of the increasingly pervasive monitoring.

7. CONCLUSION

To summarize, both the data acquisition and data mining that are used to tune personalization profiles for information filtering and the user profile models themselves are ethically contentious practices. In order to counter balance the potential privacy invasiveness of these practices they should require a high level of transparency and clearly informed consent from the service users. It is therefore all the more problematic that many users are not, or only vaguely, aware of the fact that major services, e.g. Google search and Facebook Newsfeed, employ personalized information filtering.

8. ACKNOWLEDGMENTS

This work forms part of the CaSMa project supported by ESRC grant ES/M00161X/1. For more information about the CaSMa project, see http://casma.wp.horizon.ac.uk/ .

9. REFERENCES

[1] GovLab, 2013. The GovLab Index: The Data Universe. *GovLab Blog*, (August 22, 2013), NYU Polytechnic School of Engineering. http://thegovlab.org/govlab-index-the-digital-universe/

[2] Internet live stats, 2015. Total Number of Websites. *Internet live stats*. http://www.internetlivestats.com/total-number-of-websites/

[3] Speretta, M., Gauch, S., 2005. Personalized search based on user search histories. *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on*, vol., no., pp.622,628, 19-22 Sept. 2005. doi: 10.1109/WI.2005.114

[4] Rohle, T., 2007. Desperately seeking the consumer: Personalized search engines and the commercial exploitation of user data. *First Monday*, [S.l.], sep. 2007. ISSN 13960466. http://journals.uic.edu/ojs/index.php/fm/article/view/2008/1883.

[5] Ma, H., Zhou, D., Liu, C., Lyu, M.R., King, I., 2011. Recommeder systems with social regularization, *WSDM '11 Proceedings of the fourth ACM international conference on Web search and data mining*, pp287-296. 2011. doi: 10.1145/1935826.1935877

[6] Koene, A., Perez, E., Carter, C.J., Statache, R., Adolphs, S., O'Malley, C., Rodden, T. and McAuley, D., 2015. Research Ethics and Public Trust, Preconditions for Continued Growth of Internet Mediated Research, *1st International Conference on Information System Security and Privacy (ICISSP)*, Angers, France, February 9-11, 2015.

[7] Koene, A., Adolphs, S., Perez, E., Carter, C.J., Statache, R., O'Malley, C., Rodden, T. and McAuley, D., 2015. Ethics considerations for Corpus Linguistics studies using internet resources, *Corpus Linguistics 2015*, Lancaster, UK, 21-24 July, 2015.

[8] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, 1992. Using collaborative filtering to weave information tapestry, *Commun. ACM*, 35(12), 61–70.

[9] J. Golbeck, 2006. Generating predictive movie recommendations from trust in social networks, *Trust Management, Proceedings 4th International Conference, iTrust 2006*, Pisa, Italy, 93–104, May 16-19, 2006.

[10] R. Burke, 2007. Hybrid web recommender systems, *The AdaptiveWeb*, 377–408. Springer Berlin / Heidelberg.

[11] L. Rokach, B. Shapira, and P.B. Kantor. 2011. *Recommender systems handbook*. Vol. 1. New York: Springer.

[12] Google Terms of Service, https://www.google.co.uk/intl/en/policies/terms/

[13] Google Privacy Policy, http://www.google.com/policies/privacy/

[14] A. Besmer, H. R. Lipford. 2010. Moving Beyond Untagging: Photo Privacy in a Tagged World. *CHI 2012: Privacy*, pages 1563- 1572

[15] Z. Whittaker. 2010. Facebook does not erase user-deleted content." *ZDNet, April* (2010). http://www.zdnet.com/article/facebook-does-not-erase-user-deleted-content/

[16] Netflix official blog announcement, March 12, 2010. http://blog.netflix.com/2010/03/this-is-neil-hunt-chief-product-officer.html

[17] F. Abel, Q. Gao, G.-J. Houben, and K. Tao. 2011. Analyzing user modeling on twitter for personalized news recommendations. *In User Modeling, Adaption and Personalization*, pages 1–12. Springer, 2011.

[18] N. Matthijs and F. Radlinski. 2011. Personalizing web search using long term browsing history. *In WSDM 2011*, pages 25–34.

[19] Z. Zhao, Z. Cheng, L. Hong, E.H. Chi. 2015. Improving User Topic Interest Profiles by Behavior Factorization. *In WWW 2015,* pages 1406-1416.

[20] M. Qiu, F. Zhu, and J. Jiang. 2013. It is not just what we say, but how we say them: LDA-based behavior-topic model. *In SDM*, pages 794–802.

[21] J. Tang, M. Zhang, and Q. Mei. 2013. One theme in all views: modeling consensus topics in multiple contexts. *In SIGKDD 2013*, pages 5–13.

[22] H. Yin, B. Cui, L. Chen, Z. Hu, and Z. Huang. 2014. A temporal context-aware model for user behavior modeling in social media systems. *In SIGMOD 2014*, pages 1543–1554.

[23] K. Hill. 2012. How Target Figures Out A Teen Girl Was Pregnant Before Her Father Did. Forbes 2/16/2012. http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

Cryptocurrencies as Narrative Technologies

Mark Coeckelbergh
Centre for Computing and Social Responsibility, De Montfort University
The Gateway, Leicester LE1 9BH
+44 116 257 7487
mark.coeckelbergh@dmu.ac.uk

Wessel Reijers
ADAPT Centre
Dublin City University
Glasnevin, Dublin 9
+353 87 439 51 12
wreijers@computing.dcu.ie

ABSTRACT
Transitions in monetary technologies raise novel ethical and philosophical questions. One prominent transition concerns the introduction of cryptocurrencies, which are digital currencies based on blockchain technology. Bitcoin is an example of a cryptocurrency. In this paper we discuss ethical issues raised by cryptocurrencies by conceptualising them as what we call "narrative technologies". Drawing on the work of Ricoeur and responding to the work of Searle, we elaborate on the social and linguistic dimension of money and cryptocurrencies, and explore the implications of our proposed theoretical framework for the ethics of cryptocurrencies. In particular, taking a social-narrative turn, we argue that technologies have a temporal and narrative character: that they are made sense of by means of individual and collective narratives but also themselves co-constitute those narratives and inter-human and social relations; configuring events in a meaningful temporal whole. We show how cryptocurrencies such as Bitcoin dynamically re-configure social relations and explore the consequent ethical implications.

Categories and Subject Descriptors
K4.1 [**Computers and society**]: Public policy issues – Ethics

E3 [**Data encryption**]: Public Key Cryptosystems

General Terms
Design, Human Factors, Theory

Keywords
Cryptocurrencies, Bitcoin, technology, mediation, narrative, Ricoeur, Searle, money

1. INTRODUCTION
One of the intriguing myths of our time concerns the narrative surrounding so-called "cryptocurrencies", with Bitcoin as its main instantiation [18, p12]. The technology appears to be promising: the possible applications of the underlying *block chain* technology seem to be spectacular and feasible in the near rather than in a distant science fiction-like future. According to Melanie Swift, it has the potential to bring about radical new forms of money, contracts and even governments and democracies [23].

Bitcoin's mysterious founder - or anonymous group of founders - Satoshi Nakamoto, characterises Bitcoin as an "electronic payment system based on cryptographic proof instead of trust" [16, p1]. Its architecture has been based on the underlying "'blockchain protocol', which assures transaction authenticity, integrity, and ordering" [6, p84]. Basically, the blockchain is a public ledger (like a book of accounts) that contains all the transactions made within its system. "Blocks" are records containing the most recent transactions that are cryptographically signed and added to the chain in a designated sequence, in a linear, chronological manner [23, p10].

The main innovative feature of the blockchain is not its potential for bringing about fully anonymous transactions, but its capacity to track transactions within a systems and therefore fully exclude counterfeiting [13, p33]. This feature correlates with its ability to de-centralise authority and conduct transactions on a peer-to-peer basis. In the case of cryptocurrencies, this means that governments and banks are not needed to authenticate and validate transactions; these tasks are *delegated* to the technology. Because of their capacity to challenge authority, cryptocurrencies are seen as "weapons in the new control society" [5, p7]. Transactions with cryptocurrencies are irreversible and they solidify economic contracts by turning code into economic law. Because of the great potential for social control through the technology of cryptocurrencies (social control in de-centralized form), there seems to be a significant need for developing an ethics of cryptocurrencies.

In this paper, we will address this need by creating a framework that enables us to ethically assess the implications of the crypto currency technology. We base our conceptual structure on a juxtaposition between John Searle's social ontology and Paul Ricoeur's narrative theory. We will argue that cryptocurrencies can be understood as "narrative technologies" that both configure our (economic) reality and bring about an abstraction from the practical realm of economic exchange. In accordance with this analysis, we will discuss the ethical consequences of these configurations.

2. CRYPTOCURRENCIES AS NARRATIVE TECHNOLOGIES
In this section, we will inquiry into the meaning and use of currencies by asking: what *are* they, and what *do* they? These questions need to be addressed, for no ethical implications can be derived from a phenomenon that is not properly understood and from which most possible implications lie in the future rather than in the empirical present. In order to answer these two questions, we will first of all interpret cryptocurrencies as linguistic

phenomena, whose ontology can be analysed by means of Searle's theory of social reality. We then juxtapose this conceptualisation with Ricoeur's narrative theory that will enable us to show how cryptocurrencies can configure our narrative reality.

2.1 What cryptocurrencies *are*: technologies & linguistic institutions

One straightforward answer to the ontological question of what cryptocurrencies are can be that they *are* their blockchains: that the essence of the technology is the ever-growing chain containing records of transactions. Advancing on the ontological question, we can state that the blockchain consists of *code,* of a sequence of symbols that can be read by computer algorithms. However, this "code" has a significantly human and indeed social-institutional dimension. Cryptographic code, as argued by Lessig, is similar to human-made law while it can enforce confidentiality as well as identification in similar ways as law can [14, p53]. John Searle offers an ontological grounding that explains the similarity between law and programming code by pointing at their linguistic origin. He states that all human made phenomena, ranging from streets to governments to laws, share a linguistic basis. The origin of certain artificial phenomena, called institutional facts, is traced back to linguistic entities called "status function declarations" [20]. An example of a simple status function declaration is: "I hereby declare that the provided information is true".

Status function declarations include both a locutionary aspect (a linguistic aspect) and an illocutionary aspect (an extra-linguistic aspect). They are characterised by what Searle calls a "double direction of fit", a notion which refers to the fit between the locutionary, propositional aspect of the declaration and the human directedness to the world implied by the illocutionary aspect. For declarations, two different illocutionary aspects coincide: the desire to make something the case and the belief to make something the case. In other words, if we declare something to be the case, we might *create* a reality while *desiring* it to come about. For example, when a certain person is declared to be the President of the United States, the propositional form of the declaration "I, (Barack Obama), hereby declare that…", fits with the desire to bring about a new state of affairs *and* with a new ontological reality (the new president of the United States).

If we apply Searle's explanatory model to ontologically ground the phenomenon of cryptocurrencies, we can state that they indeed are status function declarations. They are declarations because they have a propositional structure that is such that it allows them to bring about their own reality. Moreover, they are *status function*[1] declarations as their meaning depends on a coinciding structure of human *desires* and *beliefs:* when using the blockchain of a crypto currency, we belief the new state of affairs (a transaction) which coincides with our want to bring it about (we wanted the transaction to occur). These desires and beliefs, however, don't belong to the individual but to the *collective.* We can collectively *intend* for status function declarations to become part of our social reality.

However, this does not seem to be an adequate way to wholly explain the semantics of cryptocurrencies. Two main lacunas make Searle's theory incapable to serving as a solid basis for the examination of cryptocurrencies. First of all, Searle leaves the gap between individual intentions and collective intentions unexplained, stating that collective intentions are merely biologically primitive phenomena. By suggesting this reductionist view, he disqualifies the impact of *culture* that is precisely not reducible to human biology [9, p259]. Secondly, his theory does not include an aspect of normativity that is needed to explain *why* declarations can have a status function at all [9, p260]. In the case of cryptocurrencies, we would want to explain *why* we assign a status function to them. In more common terms, we would want to explain why people value cryptocurrencies. This is not a trivial point, for the meaning of cryptocurrencies (as well as their classification as money) depends on their relation to human normative values. In order to deal with the two problems of culture and normativity, we turn to a theorist who takes quite a different stance on the role of language: Paul Ricoeur.

2.2 What cryptocurrencies *do*: configuring narratives

In one of his major works, *Time and Narrative*, Ricoeur constructs a comprehensive narrative theory. Unlike Searle, Ricoeur does not focus on the formal structure of language (like the formal structure or syntax of programming code), but on its hermeneutic aspects: the way people interpret language and their life-world. His theory revolves around one basic model that describes the way in which a text considered as a narrative can mediate human reality. This model consists of three conceptual moments that indicate the move from "not having read" to "having read" a narrative. Ricoeur claims that our social reality is embedded in a *prefigured time.* This means that the way we experience our temporal, social existence is embedded in a cultural context that is shaped by narratives [17, p54]. For example, we understand ourselves due to national narratives ("I'm a citizen of the Netherlands), economic narratives ("I lost my job due to the financial crisis") and even technological narratives ("robots are going to take our jobs"). Hence, whenever we engage with language we act from this cultural basis, which means that our understanding is shaped by the narratives that are so-to-say a part of our *collective* memory.

Prefigured time indicates the moment at which we start to interact with a text. From the prefigured time, we proceed to the moment of the *configured time,* which is the backbone of Ricoeur's theory. The paradigm of configured time is the notion of the *plot* in a story. The plot is defined as an organisation of events that mediates between the heterogeneous factors (like agents, goals and interactions) and the syntagmatic order of a narrative as a whole [17, p65]. More commonly said, the plot is the organisation of elements of a story that makes it possible for a reader to follow it to a certain conclusion. This organisation depends on two distinct temporal dimensions: a chronological and an a-chronological one. The chronological dimension comes about by means of a sequence of events ("first *this* happened, secondly *this* happened"). In contrast to the chronological dimension, the a-chronological dimension makes it possible to oscillate between the story as a whole and separate events, to jump between different "times" (e.g. as happens in a flash back) and to create a sense of ending.

By means of configuration, a text *refigures* our understanding of the world we live in. The world of the text and our human world intersect during this moment [17, p71]. For our analysis of cryptocurrencies, we will focus on the *configurative* capacity of these technologies. Starting from the paradigm of a text, Ricoeur shows that the process of *configuration* encompasses two distinct

[1] A "status function", according to Searle, is the function ascribed to an entity solely because of its status. For example, the king's seal has a function because of the status assigned to it, not because of its physical properties.

capacities of narratives that are significant for our understanding of cryptocurrencies. First of all, configuration brings about an *active* process of interpretation: a narrative actively re-organises the pre-figurative understanding of a reader. An analogy with a computer process might be helpful here: in the process of *reading* data by a computer, data are simultaneously *written*. Hence, the interpretation of a narrative implies a coinciding active process of (mental) *reading* and *writing*. Secondly, narrative structures can be made increasingly abstract by means of constructing so-called second- and third order entities that are based on first order entities (real characters and events) [17, p181]. For example, socio-cultural structures like companies and countries are *abstract* entities that do not directly denote real *people* or *events.* Nonetheless, any attempt aimed at explaining these structures will include first order entities: it will include narratives about real people who act in real situations.

Unlike Searle, Ricoeur addresses the two aspects of linguistic mediation of social reality we discussed in the previous section. Firstly, he characterises narratives as *cultural* phenomena: we interact with narratives from within our cultural embedded (prefigured) situation. Secondly, he explains *why* narratives can configure our social reality: they have the function of emplotment. Emplotment has an outspoken *normative* character because the characters in a narrative are not just "doers" as Searle would portray them but are "endowed with ethical qualities" [17, p59]. Unlike generalised "doers" like the *homo economicus* who act mechanically according to non-normative motives*,* characters can be good or evil; the protagonists or antagonists of their (life-) stories. These two features of Ricoeur's theory enable him to go beyond Searle's formal approach and to provide a holistic and normative account of linguistic mediation of our social world.

How then, could we employ Ricoeur's narrative theory to understand the *technological* phenomena of cryptocurrencies? We want to explore in what sense technologies could be "narrative". But since Ricoeur's theory revolves around the paradigm of the text, both as history and as fiction, we need to justify the claim that the concept of a narrative in a text can be extended to the concept of a narrative *technology*. Can Ricoeur tell us something about technology? Technology only plays a very marginal role in Ricoeur's work. However, Kaplan has drawn a connection between Ricoeur's work and the philosophy of technology. He suggests that Ricoeur's hermeneutical method as well as his analysis of the hermeneutic circle between human experience and narration can be fruitful in discussions about technology [11, p43-44]. Moreover, he argues that "the model of the text is also the model for the mediation of experience by technology" [11, p169]. Thus, Ricoeur's theory can be used to improve our understanding of technology.

Our conceptual model of narrative technologies is inspired by Ricoeur's model of emplotment. We argue that technologies configure our narrative understanding by organising events into a meaningful whole that includes both humans and things. For instance, a *car*, as a technology, configures events like "starting the engine" and "adjusting the mirrors" in a meaningful whole that includes both human and non-human *characters*. However, technologies do not configure our narrative understanding in only one single way for some of them might be very similar to the paradigm of the text and others very different. Nevertheless, we argue that these differences are matters of *degree* rather than matters of differences in *kind.* All technologies affect our narrative understanding, but the extent to which this is the case and the ways in which they do so differs between technologies. We propose two distinctions: one between active and passive narrative technologies, and one between abstracting and engaging narrative technologies. Let us explain these distinctions and use them to develop our hermeneutic framework.

The first distinction relates to the capacity of technologies to constitute an active process of interpretation. The degree of activity is determined by the extent to which a technology closes in on the paradigm of the text. Some technologies have very little in common with the paradigm of the text and for the most part play a role in our prefigured understanding. For instance, a bridge is part of a prefigured narrative structure in which events and characters are already configured into a plot: for example it may be a bridge to transport goods and people across the Rhine river. When a bridge gets built, it plays a role in configuring our narrative understanding (for example by disclosing new areas of a country) but soon it becomes part of our prefigured time; a *passive* element of our narrative understanding. However, some technologies *actively* configure our narrative understanding. They can "read" and "write" our narrative understanding by means of emplotment. ICT technologies are exemplary for this type of narrative technologies and are most similar to the paradigm of the text. This can first of all be derived from their very "textual" character: many forms of human-computer interaction revolve around mediation by symbolic and textual information. More importantly, though, ICT technologies and humans so-to-say "co-author" and "co-act" the narratives they engage in. Consider for instance video games. Players can interact with each other in a game, which *also* organises the characters and events into a plot. The unfolding of the narrative is co-created by the technology and the humans. The technology can be explicitly *narrative* and *social*.

The second distinction we propose is one between abstracting and engaging narrative technologies. Technologies have the capacity to create distance, which can be understood in two ways: as creating a distance between people and between people and things. In line with Ricoeur's theory, we argue that abstracting technologies remove themselves from the realm of action by configuring quasi-characters and quasi-events in a plot. Monetary technologies bring together not so much humans and direct interactions between them, but rather quasi-characters and quasi-events; also referred to as "second- and third-order" entities by Ricoeur as opposed to "first-order" entities which are actual characters and events. They organise quasi-characters such as "markets" and "exchanges" and quasi-events (e.g. algorithmic trades) in quasi-plots (e.g. the flash crash)[2]. This interpretation of what monetary technologies do is also in line with the claim – partly inspired by Simmel [22] – that modern financial technologies have abstracting effects Modern technologies even render time itself abstract, as Ricoeur suggests: the machines that measure time enable an: 'abstract representation of time" [17, p63]. Indeed, modern time technologies (clocks) serve to abstract time from concrete events, characters, and plots. Similarly, one could argue that the configuration of our narrative understanding by Fordist production technologies such as the conveyer belt abstract from the narrative of engaging labour (the artisan). Thus, these technologies distance themselves from the direct narrative they constitute, from the organisation of events in which actual

[2] Heidegger discussed such a process of abstraction as a typical feature of modern technology [8]. A hydroelectric plant configures people and the electricity they use in a way that abstracts from actual characters and events.

characters play a role. Engaging technologies, by contrast, instantiate a narrative as a direct interaction between human and non-human characters in actual events. Instances of such engaging technologies can be pre-modern ones like hammers, but not exclusively so. Modern ICTs can likewise create engaging narratives that re-situate people as characters that can partake in a narrative. Video games are primary examples of ICTs that enable engaging narratives but online communities like Github or Wikipedia can be said to do the same. These kinds of technologies engage people as actual characters in the plot of a digital narrative.

The above analysis gives us four categories for a hermeneutics of narrative technologies, with crypto-currencies assigned to one of the cells of the matrix: the category of active and abstracting narrative technologies:

Table 1. Hermeneutics of narrative technologies matrix

Narrative technologies:	**Abstracting**	**Engaging**
Passive	Power plant	Bridge
Active	Crypto-currencies	Video games

In line with the above-mentioned schema, we can now give a more precise description and understanding of what crypto-currencies are and do: we argue that crypto-currencies are active narrative technologies that abstract from the narrative they instantiate. First, they are active as they time-stamp transactions and thus co-create the transaction narrative. Through the technology, the human event of a transaction becomes an integral part of computational bookkeeping. The technology thus configures what we may call an "accounting" narrative. Second, although cryptocurrencies mediate events (transactions) between actual characters (traders, consumers), they remain remote from that level of human action and instead operate on a calculative, mathematical level. The transactions become a matter of algorithmic calculations; they are removed from actual people and events – including from the concrete material realities such as the goods that are traded and from the computing infrastructure. Hence, they can be said to create distances, both between people and between people and things.

3. THE ETHICS OF CRYPTOCURRENCIES

What does this mean for the ethics of cryptocurrencies? Usually, computer ethics but also ethics of finance are concerned with values such as privacy, democracy, autonomy, and with the behaviour of humans such as bankers, money traders, etc. and the fairness of financial institutions. For instance, Boatright [1] sees finance ethics as being concerned with the fairness of markets and the duties and rights of the participants in those markets. Technology is considered, but is seen as normatively neutral, or is view in a merely instrumental or consequentialist way, for instance by asking: do cryptocurrencies enable fraud? Do miners act responsibly? Do cryptocurrencies lead to more democratic economic and political systems? These questions are important, but lack a connection with humans as narrative beings that understand their world through an interaction with technologies. By contrast, this paper proposes a different approach that may complement existing approaches in ethics of technology: the focus is on tying to understand how technologies configure our narrative understanding. We offer a framework that enables us to analyse the narrative hermeneutics of financial technologies such as cryptocurrencies. Let us now say more about normative dimension of this narrative-hermeneutical role and the ethical implications these technologies bring about. We propose to structure this discussion by distinguishing between the configuring and the abstracting functions of crypto-currencies.

First, we consider the way cryptocurrencies influence our narrative understanding both *passively* (as elements of our prefigured narrative understanding) and *actively* (as technologies that actively configure the understanding of characters they interact with). These processes are not normatively neutral. The notions of *transaction* and *trust* are central to the rationale of cryptocurrencies, and the way transactions and trust change through the new technology has ethical implications. Let us briefly discuss these normative aspects and ethical implications of crypto-currencies as configuring narrative technologies.

When analysing the *prefigured* time in which cryptocurrencies play a role, we have to consider the normative-ethical dimension of the narrative structures that surround money as a technology. Cameron argues that our understanding of the monetary system is thoroughly shaped by narratives. Recently, these narratives have been placed in the greater context of the global financial crisis. This is everything but ethically neutral. Cameron forcefully shows how abstract financial processes are broken down into narratives about people (bankers, traders) that are *characterised* as "Gods" and "demons" [2, p12]. Systems that were perceived as being ruled by abstract rational calculations appeared to be embedded in a narrative structure incorporating characters with strong ethical qualities. The wake of cryptocurrencies can be interpreted in line with these global economic and political narratives. One of the major catalysing factors in the development of Bitcoin was the political blockade of Wikileaks by the world's major payment companies [18]. On the one hand, this blockade revealed the narrative structure containing the roles these companies play, which showed that the assumed neutrality of the monetary system was illusory. On the other hand, the emergence of Bitcoin configured this narrative understanding by presenting an alternative based on two distinct features: the decentralisation of power and the delegation of *trust* from legal authorities to the authority of the blockchain protocol. The emerging narrative is one of securing the integrity of the monetary system independently of authorities whose supposed neutrality was shown to be ill founded. Cryptocurrencies are thus part of a normative-ethical field where different narratives compete.

However, we have shown that cryptocurrencies are not merely passive elements of our prefigured understanding, but actively configure our narrative understanding through human interaction. Let us focus on the notion of *transaction* to show how this configuration takes place and what ethical consequences it brings about. A transaction may be defined as a configuration of human action (acting *through* something, as the term implies). Georg Simmel, in *The Philosophy of Money,* shows how the development of money has transformed transactions (which he grounds in the notion of *exchange*) from more direct forms to less direct, abstract forms [22]. We can reframe Simmel's theory by using the conceptual apparatus developed earlier in this paper. The more direct, original form of inter-human exchange has a narrative character in the sense that it configures events (deliberating on a price, handing over goods) between characters (the merchant, the farmer) in a meaningful whole we call a transaction. But with the introduction of money, this configuration changes. Money, according to Simmel, mediates these transactions and makes them indirect. People now interact

through money to engage in the narrative structure of exchange. Monetary technologies, as active narrative technologies, therefore configure our understanding of transactions: they no longer mediate organisations of events "between people" but actively configure these organisations through the use of a technology. This different kind of configuration is not normatively neutral: it concerns the way we think about and construct what transactions *should* be. But then how do cryptocurrencies configure our understanding of transactions? Based on the previous analysis, we could say that they render the a-chronological dimension of the transaction narrative obsolete by enforcing chronological time (time-stamped transactions) in their systems. This configures the understanding of making a transaction from an organisation of events with no fixed order that can be reversed, to one with a fixed order that is irreversible. This has an ethical implication, as it constrains the transaction in a specific way. Therefore, we have to decide and discuss whether this is what we want transactions to be, or if want another configuration.

Secondly, cryptocurrencies create a distance between the narrative structure of economic exchange and the transaction process contained in the blockchain protocol. Again this is a normative-ethically significant shift in our understanding of exchange and transactions. Simmel already questions the processes of abstraction and distancing entailed in the development of modern money [22, 4]. With crypto-currencies, this modern process of abstracting and distancing now further increases. Mediated through the blockchain technology, economic exchanges and financial transactions seem now even more abstracted from concrete people and events. Transaction now seems entirely a matter of numbers and algorithms. To say it in a Simmelian way: the quantification of modern life seems now to have reached a (new?) summit. Again, we may discuss whether this quantification is desirable and acceptable and draw consequences for the financial technologies we use.

This normative shift is also illustrated by the notion of *trust* as it is used in the established rationale of cryptocurrencies. The trust between the first-order entities in the narrative, the miners, the traders and the cryptocurrency users, is substituted by the systemic rigidity of the technology, that is, by a second-order entity. People using a cryptocurrency know that they are dealing with authentic and validated transactions not because they can trust the other people in the network, but because these features are *enforced* in the system. However, it would be a mistake to suppose that the notion of trust is altogether removed from the use of the technology: trust is still needed [4]. However, rather than trusting the people in the system, we need to trust the system itself. In modern times, trust related to first-order entities such as persons and material forms of money was already replaced by trust in a more abstract monetary system (think for instance about trust in "the system" after the end of the gold standard). As money dematerialized, trust also already depended increasingly on what was written down and recorded [4]. But now blockchain technology seems to further enhance this function of money by turning trust between people into trust in technology and systems. Just as a text might abstract from a narrative by substituting first-order with second-order entities, cryptocurrencies can be said to similarly do so, as they create (new) second-order entities, for instance Bitcoin technology and Bitcoin currency, which may be trusted (or not), depending in turn on changes in our narrative understanding. Explaining these second-order entities will always involve a referral to first-order entities. For example, whenever a cryptocurrency system gets affected by malfunctioning software, by attacks on cryptocurrency exchanges or by any other intended or unintended factor, the narrative structure of first-order entities (miners, hackers, programmers) is revealed. Apparently abstract entities such as "Bitcoin" still depend on concrete people and events. However, usually these are hidden from view. We only see them when the technology breaks down, when trust is (already) eroding. Moreover, although there is 'trust in the algorithm' (or not) [4, p165], the technology still requires 'trust between people': if no-one trusted and *used* Bitcoin, for instance, then it would not work; trust now also depends on whether our peers use it – revealing that transactions and trust in the financial sphere and elsewhere were always already a *social* matter [4, p165].[3]

What further ethical implications can be derived from this analysis? First, it leads us to a discussion on the ethics of transactions. By allowing transactions to be delegated to blockchain technologies, and therefore by getting rid of the a-chronological dimension of inter-human exchange, we are able to transform social relations (including contractual relations) in fully rigid forms. Now one could argue that for some social relations such as financial transactions, this level of rigidity can be very beneficial for it prevents cases of fraud, counterfeiting and "creative bookkeeping". However, this draws a firm line between on the one hand those areas of human exchange that we want to fully formalise and configure like crypto-currencies do and, on the other hand, those areas of exchange that would benefit from the human freedom implied in the a-chronological dimensions of exchange. Is such a clear boundary conceptually and practically sustainable? In any case, if we draw such a line we can then argue that there are illegitimate boundary crossings between two spheres. For instance, we may want to prohibit a crossing from the financial sphere to the sphere of health care. When informal human relations, like the relation between caregiver and caretaker, would be put in the rigid format provided by cryptocurrencies, inter-human relations might become "entangled" in their technological dependency as argued also by Hodder [10]. We would regard the transaction or the contract as the end-points of our relations with other human beings, rather than intermediate relay stations. In case a contract is practically breached, the blockchain protocol itself will be the arbiter: its acceptation or rejection of a transaction functions as the final verdict without a question being asked as to whether the transaction is ethically desirable in the first place. For example, in the context of care relations, a blockchain approach would mean that those relations do not only become very contractual and impersonal, but also that there is no room for interpretation and revision. Assuming this might be undesirable, we may want to avoid and prevent this from happening.

Secondly, our narrative approach can be used to reveal the ethical consequences of the abstracting narrative capacity of cryptocurrencies. As Simmel shows, these consequences can be both positive and negative. The positive effects of abstracting monetary technologies like cryptocurrencies lie in their capacity to emancipate and empower people. If social relations become less personal, then this also renders them more free: relations become a matter of choice and money becomes a guarantee of inclusion in the realm of economic exchange, regardless of your

[3] Consider also Heidegger's distinction between present-at-hand and ready-to-hand [7]: we could say that first-order entities are revealed when, because of malfunctioning, the technology appears to us present-at-hand rather than ready-to-hand.

personal, racial, or cultural background and status[4]. Moreover, like other forms of money the technology enables you to communicate and transact with anyone on earth; there are no physical-geographical boundaries. These effects are in line with the predominantly cosmopolitan and libertarian ideology most present in the narratives offered by cryptocurrency communities [12]. Since within the system any transaction can be performed without the parties involved in the transaction having to trust one-another, *who* you are is totally irrelevant. This enables people from any kind of background to engage in transactions without an authority preventing them to do so. Moreover, it is said that cryptocurrencies could empower people to gain the benefits from financial services in developing countries that have so far been secluded from access to banking services [3]. Crypto-currencies thus seem to radiate positive and optimistic ethical and political promises. However, as Simmel already suggests, the abstraction from the narrative of inter-human exchange comes with a cost. Firstly, by delegating the trust in transactions from first-order entities to second-order entities, the responsibilities of people acting through the system are delegated to the level of the system itself – thereby excluding and rendering invisible the level of first-order entities and events. Whatever kind of transaction one performs through the system, the only normative check is whether the system allows or declines it. What kind of transaction is performed (which can be a "good" or a "bad" transaction) is irrelevant. Currently, these technological loopholes have to be countered by legal measures and as yet it is unsure how this can be dealt with in the future. Furthermore, politically speaking the abstracting capacity of cryptocurrencies will likely have significant effects on power-relations between people and institutions. With *trust* being delegated to the second-order entity of the cryptocurrency system, power struggles might arise; first between cryptocurrencies and states (several states like Iceland and China have fully or partly banned cryptocurrencies) but more importantly *between* cryptocurrencies. While already banks are investing huge sums of money in blockchain technology [19] and cryptocurrencies might be viable forms of state currencies [15], it is uncertain that the decentralised features of the technology will also result in decentralisation of institutional power. Since the ability for social control is optimised within a cryptocurrency system, the question of who controls the system remains of pivotal ethical and political importance. Hence, to describe, reveal and discuss these political implications we are again forced to explain the changes at the systemic level of second-order entities by means of a narrative of first-order entities: of real characters interacting through an organisation of events.

4. CONCLUSION AND IMPLICATIONS FOR THE ETHICS OF CRYPTOCURRENCIES

"All the world's a stage, and all the men and women merely players" [21, p52]. These words of Shakespeare remind us of the importance to consider the impact technologies have on the narratives that shape our lives. We are narrative beings; yet we are also technological beings, and in contrast to what many people may suppose, both are related. In this article we have argued that technologies are not merely "narrative" in the sense that they are part of the narrative we – as persons, communities, societies, and cultures – tell about ourselves; technologies do much more: they also shape these narratives. As our analysis shows, financial technologies are no exception, and as they actively configure our understanding of financial practices and abstract transactions between people, they re-shape human and social reality in normatively significant ways. If we want to discuss the ethics and politics of finance, therefore, it is important to take into consideration these financial technologies and their narrative capacities. For crypto-currencies, this means that discussing finance in the light of these new technologies requires us to attend to their ethical and political implications as narrative technologies. We have argued that crypto-currencies do and might change transactions, trust, and power – and indeed the very way we think about these concepts. Discussing about crypto-currencies, therefore, is not a "technical" matter but does and should concern all of us. It is not so much about "ethical issues" with cryptocurrencies but about how we might re-imagine and re-design the social – how we might tell new, better stories about ourselves and indeed stage a better play: one we think is more ethically and politically responsive and responsible. Ethics of finance is, of course, about people. But *therefore* it is about the technologies-narratives we want.

[4] See also [22, p324]

5. ACKNOWLEDGMENTS

The ADAPT Centre for Digital Content Technology is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

6. REFERENCES

[1] Boatright, John R. (ed.). 2010. *Finance Ethics: Critical Issues in Financial Theory and Practice*. Hoboken, New Jersey: John Wiley & Sons.

[2] Cameron, A. 2015. Money's unholy trinity: Devil, trickster, fool. *Culture and Organization*, (May 2015), 1–16. http://doi.org/10.1080/14759551.2015.1035721

[3] Clegg, A. G. 2014. Could Bitcoin be a financial solution for developing economies? *University of Birmingham*.

[4] Coeckelbergh, M. 2015. *Money Machines: Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. Farnham: Ashgate.

[5] DuPont, Q. 2014. The Politics of Cryptography: Bitcoin and The Ordering Machines. *Journal of Peer Production*, *1*, 1–10.

[6] Folkinshteyn, D. 2015. A tale of twin tech: Bitcoin and the www. *Journal of Strategic and International Studies*, *X*(2), 82–90.

[7] Heidegger, M. 1927. *Being and Time* (trans. J. Stambaugh). Albany, NY: SUNY Press, 1996.

[8] Heidegger, M. 1954. The Question Concerning Technlogy. In *The Question Concerning Technology and Other Essays* (trans. William Lovitt), New York: Harper & Row, 3–35.

[9] Heidemann, C. 1999. On Some Difficulties Concerning John Searle's Notion of an "Institutional Fact." *Analyse & Kritik*, *20*, 143–158.

[10] Hodder, I. 2014. The Entanglements of Humans and Things: A Long-Term View. *New Literary History*, *45*(1), 19–36. http://doi.org/10.1353/nlh.2014.0005

[11] Kaplan, D. M. 2003. *Ricoeur's Critical Theory*. New York: State University of New York Press.

[12] Karlstrøm, H. 2014. Do libertarians dream of electric coins? The material embeddedness of Bitcoin. *Distinktion: Scandinavian Journal of Social Theory*, *15*(1), 23–36. http://doi.org/10.1080/1600910X.2013.870083

[13] Kostakis, V., & Giotitsas, C. 2015. The (A)political economy of bitcoin. In *P2P & inov.,* (Rio de Janeiro, 2015) Vol. 2, No. 2, 28-44

[14] Lessig, L. 2006. *CODE version 2.0*. New York: Basic Books.

[15] Malefijt, L. de W. 2014. *NLCoin.* Bachelor Thesis. Universiteit Utrecht.

[16] Nakamoto, S. 2009. *Bitcoin : A Peer-to-Peer Electronic Cash System.* (White paper).

[17] Ricoeur, P. 1983. *Time and Narrative - volume 1* (Vol. 91). Chicago: The University of Chicago Press. http://doi.org/10.2307/1864383

[18] Roio, D. J. 2013. Bitcoin, the end of the Taboo on Money. *Dyne.org Digital Press*, (April 2013), 1–17.

[19] Samman, G. 2015. Blockchain Tech is the "Biggest Opportunity" for Banks to Stay Relevant, Says New Report. Retrieved June 21, 2015, from Cointelegraph: http://cointelegraph.com/news/113805/blockchain-tech-is-the-biggest-opportunity-for-banks-to-stay-relevant

[20] Searle, J. R. 1995. *The Construction of Social Reality*. London: Penguin Group.

[21] Shakespeare, W. 2005. *As you like it*. San Diego: ICON Group International. Retrieved from http://eprints.kingston.ac.uk/1511/

[22] Simmel, G. 1900. *The Philosophy of Money*. (D. Frisby, Ed.) (3rd ed.). New York: Routledge Classics, 1978.

[23] Swan, M. 2015. *Blockchain: Blueprint for a new economy* (first edit). Sebastopol, CA: O'Reilly Media.

The Ethics of Driverless Cars

Neil McBride
Centre for Computing and Social Responsibility
De Montfort University,
The Gateway, Leicester
LE1 9BH
Email: nkm@.dmu.ac.uk

ABSTRACT

This paper critiques the idea of full autonomy, as illustrated by Oxford University's Robotcar. A fully autonomous driverless car relies on no external inputs, including GPS and solely learns from its environment using learning algorithms. These cars decide when they drive, learn from human drivers and bid for insurance in real time. Full autonomy is pitched as a good end in itself, fixing human inadequacies and creating safety and certainty by the elimination of human involvement. Using the ACTIVE ethics framework, an ethical response to the fully autonomous driverless cars is developed by addressing autonomy, community, transparency, identity, value and empathy. I suggest that the pursuit of full autonomy does not recognise the essential importance of interdependencies between humans and machines. The removal of human involvement should require the driverless car to be more connected with its environment, drawing all the information it can from infrastructure, internet and other road users. This requires a systemic view, which addresses systems and relationships, which recognises the place of driverless cars in a connected system, which is open to the study of complex relationships, both networked and hierarchical.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues: Ethics

General Terms

Human Factors

Keywords

Driverless cars, ethics, full autonomy.

1. INTRODUCTION

As traffic congestion increases, pollution from traffic, particularly in cities becomes more of an issue and accidents caused by cars kills millions, there is a need for changing how transport is managed and applying ICT and intelligent systems to providing and managing transport. This requires many systems, connecting networks of systems, understanding how communities work, developing and managing systems within smart cities. Smart transport is community based. It relies on a balance of autonomy, cooperation and regulation. It involves an exchange of knowledge between experts and users; linked systems, addressing choice, flexibility and safety. It involves cooperation with hierarchical systems without compromising freedom and democratic rights. It involves quantitative analysis combined with qualitative perception of people's attitudes, their fear and worries, and an empathic listening to the needs of transportation users.

One component of smart transportation will be driverless cars, operating within cooperative networks, relieving drivers of effort, communicating with each other, and taking into account context and conditions. Interest and investment in connected and driverless cars has grown rapidly. For example, the UK Government are investing £200 million in research and development into driverless cars. A KPMG report suggested that connected and driverless cars will create 320,000 jobs in the UK. John Leech, Head of Automotive for KPMG, commented that: "Connected and driverless cars will reduce pollution, save lives and promote social inclusion. We owe it to everyone to make this future a reality, " [4}.

The technology for autonomous cars, which drive themselves is developing rapidly. Driverless vehicles could rapidly become common sights on road. Such driverless cars will internally integrate many different inputs to analyse their environments and issues such as road conditions in order to make decisions about where and when to drive. Using learning algorithms, they will build knowledge bases of road conditions and learn to managing unusual and exceptional conditions such as plastic bags blowing across the road, or obstructions in the road [7].

As well as having integrated systems and sensors internally, they may draw on GPS to navigate, or communicate with road traffic systems such as traffic lights, and with other autonomous vehicles, to make decisions about overtaking and to form road train. Autonomous cars may report standard data to central hubs, rather like an aircraft sends out standard data. Location data will enable it to plan routes and change routes as it is progressing. These robot cars will be able to manage

Conference'10, Month 1–2, 2010, City, State, Country.

traffic flow through communication and effectively self-organise traffic systems. Both vehicle to vehicle (V2V) and Vehicle to infrastructure (V2I) may become standard parts of autonomous cars.

The networked, driverless car could be centrally controlled, delivered to a specific location and then programmed to take a fixed route. It can depend on GPS navigation, it can be part of a transport infrastructure, subject to control signals from the infrastructure. It could be constantly monitored and open to intervention from traffic controllers and police, for example, as well as the driver.

These cars may decide when driving conditions are unsuitable. They will learn to improve at driving; they will bid for insurance in real time. Where two cars are involved in a collision, active buckling control may activated so that the cars share information in order to determine who take the brunt of the collision, who buckles most based on size of the car and other factors. The driverless car will offer the driver freedom to sit back, or complete work and to save time. Driving may become a thing of the past. Commuting may become a chance to hold meetings, finish documents.

While many commentators refer to connected cars, some projects are pursuing the development of autonomous driverless cars which are self-contained and have no dependency on anything external to their own capability. These cars will act without a need to consult controllers or satellite navigation systems, independent of infrastructure, not requiring to communicate with other cars, whether driverless or not. All the information the car uses to drive it has learnt for itself. It is an individual, standing alone, driving to its own agenda, neither transmitting information nor needing to receive information. Such a truly autonomous car is a self-contained system, relying on the information it derives from its own sensors and the learning algorithms built into its on-board computer systems.

Such cars offer the possibility of complete autonomy for the car, which is in effect a robot which can navigate and make its own decisions about when and how to drive as well as how to get to a location. The ambition for autonomous cars is a complete autonomy which does not rely on infrastructure, connection with central systems or even with GPS, an autonomy where the human is eliminated from the loop, an autonomy where the car is self-determining, self-correcting, eventually self-healing and perhaps ultimately self-aware.

This ultimate autonomy may not only prove to be undesirable, but ultimately unachievable. Perhaps born of a philosophy which sees the individual as the paramount object of focus. The truly autonomous car is an enlightened car truly able to think for itself, to employ and rely on its own capabilities to determine what to believe about its environment and how to act. Technological advances and a reliance on the firm ground of mathematics will release it from its self-incurred immaturity; from its inability to use its own understanding without the guidance of another.

Using Oxford University's Robotcar as an example, this paper critiques the philosophy behind autonomous cars with the intention of working towards an inclusive, community ethics which recognises that the deployment and use of autonomous cars is with the context of community and culture and should not be considered as a isolate individual ethics outside the system and relationships within which it operates.

Firstly, the properties of Oxford University's RobotCar are examined and some ethical issues raised. The philosophy and assumptions behind the Robotcar are critiqued with reference to a talk by Paul Newman, the leader of Oxford University's programme in an Intelligence2 debate in 2013 [5]. Finally, the concepts of an autonomous car and the ethics are considered using the ACTIVE ethics framework and conclusions are drawn.

While autonomous cars offer great benefits , it is important to recognise the limits of self-reliance and that the role of an autonomous car should be considered in the context of the complex social systems and communities within which it operates.

2. OXFORD UNIVERSITY'S ROBOTCAR

Robotcar is a modified Nissan Leaf with cameras and laser sensors. It has fly-by-wire control for all aspects including steering wheel, indicators and brakes. It uses a pair of stereo cameras to navigate and lasers to assess 3D structure of the environment. It will stop if a pedestrian walks in front of it.

> *Oxford's autonomous vehicle tech does not rely on GPS to navigate because GPS doesn't work well in built-up environments and is in any case not precise or reliable enough to give an exact location. Neither does the Oxford approach use embedded infrastructure such as beacons and guide wires, which often guides robots in factories, as this would be impractical and far too expensive for use in most environments.*
>
> *Instead the Oxford approach to navigation uses algorithms that combine machine learning and probabilistic inference to build up maps of the world around it using data form on-board sensors and 'learn as it drives'. The maps it builds (and updates) are like memories of a route which can be accessed to allow the vehicle to guide itself through places it has been before.[8].*

A vehicle controlling computer runs the car, in collaboration with an interface computer and an iPAD. The autonomy is based on a localisation system which uses probability and estimation algorithms to learn about its driver and the environment it is working in. It builds up memories of past historic experiences to refer to. This experience-based navigation uses graphs (path memories) to represent the experience and various algorithms to optimise the time-

constrained localisation and overcome the problem that as the robot's memories build up, processing takes longer and the robot becomes slower to react.

Hence Robotcar is almost totally self-reliant (although it can carry out on-the-fly internet queries.) This is the goal of total autonomy. In the Robotcar case, the ambition is using available technology to create "£500 autonomy"

3. THE PHILOSOPHY OF ROBOTCAR

What is the philosophy behind Robotcar? What are the goals, the telos of its developers? The head of Robotcar, Professor Paul Newman has given many talks on the use of modern robotics in smart transport, particularly one at a debate on Smarter Mobility in 2013 [5]. I will comment on this talk and other inputs from Newman.

The rhetoric of the Robotcar is one of machine as a response to human inadequacy. Cars cause congestion, accidents and time wasting. They are inefficient. They maim and destroy. This is not because of the machine itself, but the humans who are inadequate. Robotcar involves the elimination of humans from driving because 'machines are better at doing stuff than humans." Robotics will fix human inadequacy by replacing humans: Let the cars drive; let the cars bid for insurance in real time; let the cars decide when they can drive; Let the cars get better at driving over time.

Newman draws on a range of reasons why driverless cars are important, particularly with infrastructure-free navigation and full or restricted full autonomy [1]. Firstly there is the rhetoric of safety. Safety is often used as a reason from reducing human interaction and promoting machines. This has been a strong argument for personal health monitors and controlling smart houses. Safety may mean that the designers take over and create functionality which may not be in the user's interest, or required by the user. For example, Google's car intentionally goes over the speed limit to match expected behaviour of other road users. If not known about, this may be deceptive. Hence safety is used as a reason for imposing more control on the human. Secondly, it can be argued that humans are enslaved by cars and that driverless cars free the user. Setting the slave free. Equally it could be argued that driverless cars create a new slavery. The loss of control and interaction is disempowering, deskilling the driver, leaving the driver without knowledge, skills, or the ability to navigate. The occupant of the driverless car, formerly a driver or passenger becomes an object to be moved around from place to place, controlled, moved from A to B efficiently.

Having a goal of full autonomy assumes that self-reliance is good. With no infrastructure, no reliance on authorities, road systems, or other drivers this is real autonomy, an autonomy in which the individual stands alone, independent, the sole arbiter. This may be a working out of an enlightenment view that everyone makes their own decisions without reference to others, without negotiation and outside society or communities. The car will offer the driver autonomy. But who has the autonomy, and who has had control taken from them? This autonomy could be interpreted as a loss of autonomy, a handing of control to the machine. It is the machine that is autonomous and the human that is disempowered. Autonomy is viewed black and white rather than a progression, a negotiation of interdependency between the supporting machine and the supported human.

We also see the rhetoric of "saving precious time". Time is seen as a commodity. We do not ask about the value of that time. Do we replace driving by playing video games? Saving time as a quantitative good may not mean we are gaining anything of human and moral value. And may not reference relationships. For example, driving with my son creates a non-threatening environment where he may open up and we have useful discussions.

Newman revels in the possibility of an autonomy arms race. Autonomous functions will evolve as manufacturers develop new functions in cars. Behind this perhaps is a war? A battle for power and control between the machine and the human?

Newman treats the autonomy of cars as an inevitable outcome, a sole good and a philosophical end in itself. Indeed, in his talk he pitched this as a belief: "If you don't believe this you need to leave .. this has to be a true thing."

4. THE TECHNOLOGICAL UTOPIA AND AUTONOMOUS CARS

The utopian position for autonomous cars is one that removes any reliance or connection with outside humans or technology. Even the use of satellite navigation is frowned upon. The ultimate robotcar will rely on the supremacy of the algorithm, as the sole source of truth, a truth that is amenable to logical analysis and proof.

Technology is seen as invincible, provable, permanent, materially-grounded, and reliant only on the solidity of physical laws and mathematics. It is clean, amoral, invulnerable, repeatable, unstained. The only threat of compromise and failure comes from humans included in the loop. Therefore our ultimate goal is the complete exclusion of the humans and the full autonomy of the technology.

The technological utopia contrast the reliability with the fallibility of the human. They are vulnerable, flawed, pathetically unreliable and dangerous. They harbour messiness, uncertainty, unpredictable emotions. Their unpredictability defies mathematical modelling. They cannot be trusted, freed, allocated responsibility because they will inevitably mess things up.
In the technological utopia, the technology absorbs the human. And where problems occur they point out the inadequacy of partial autonomy. Full autonomy does not recognise the interdependency of the machine and the human.

5. AN ACTIVE ETHICS RESPONSE

I will examine an ethical response to Robotcar using the ACTIVE ethics framework [6].

5.1 Autonomy

For autonomous cars, an important concern is the balance of control between the human and the robot car and the negotiation and transfer of that control based on environmental conditions. There is never total autonomy, the robotcar is an artefact encoding the perceptions of the environment held by the robot engineers. Power and control may be handed to the algorithm developers who decide the rules encoded in the programme. Autonomy is always balanced across an interface between the human and machine.

The goal of total autonomy may be a goal of deontological disconnection, and ethics in which relationship disappears and the individual is the ultimate arbitrator.

Total autonomy seeks to remove all threats by eliminating any dependency on the outside. If I am self-contained, I'm invulnerable. I have the inner certainty of provable reliable algorithms, rationally supported by the laws of mathematics. Uncertainty, risk and failure is eliminated. I do not even risk a dependency on GPS, which is considered unreliable in built-up areas. The messiness and unpredictability of the human is eliminated from the system. Total autonomy removes the human from the loop: I am logically invulnerable.

And yet the very act of moving on a road in a material world will create vulnerability and risk. My model of a rational, controlled world breaks down the moment I interact with it. There is a leakage into the machine which compromises the total autonomy.

If I elect to learn from the world I only accelerate the leakage by importing unpredictability and messiness into my autonomous system. Robotcar will be expected to learn the style and culture of driving of its host. It could well learn reckless and bad driving. If Robotcar learns the driving and cultural style of its host it simply reproduces the irrationality of the host and renders the pursuit of autonomy pointless/

So total autonomy is sterile, only actioned by deciding to eliminate interaction with the human world and minimise or restrict interaction with the material world. The autonomous system is soon rendered helpless and impotent since any interaction with the surrounding environment involved uncertainty and unpredictability and compromises autonomy. The range of options must narrow and narrow until the only reliable interaction for an autonomous system is with itself or another autonomous system which exactly reproduces its behaviour and is therefore completely predictable. All behaviour outside the autonomous system which does not conform to expectations must be ignores, discarded or eliminated.

It seems to me that total autonomy is not only unachievable but undesirable. The ambition of autonomous systems developers should be focussed on the human / robot interface, the exchanges at the interface and the balance of autonomy and control.

The quality and acceptability of a driverless car may depend not on its withdrawal from the environment and its self-containment, but on the capability of the car to interact with its environment and the richness and depth of the interactions which take place with its environment. For the developer, the focus should be on the development of communication, and the capability of rich interaction with the natural, technical and infrastructure systems.

5.2 Community

Driverless cars are created out of the interactions of a community, supported by a community of workers and serve a community. They are elements of a community, both as a participants in a relationship between humans and technology and as a technological mediators in social relationships. These cars will depend on a wide range of human interactions and human systems. Communities of cleaners, mechanics, managers, monitors and surveillance staff will support their day-to-day running. Large supplier chains will provide parts, servicing, training for maintainers and regulators. Energy suppliers and the public servants regulating practice will play significant roles. The driverless cars will be extremely dependent of the human communities which will be required to put them on the road and keep them there.

Autonomous cars will exist in a dynamic community. The pedestrians, shop keepers, police, traffic wardens will all interact with driverless cars. The technology connects the community. It cannot exist in desert-island-like isolation. Without community, we can end up with a private transportation hell, where totally autonomous cars compete for parking spaces and clog city centres. Community is about negotiation and compromise: human to human, human to machine, machine to machine.

The suspension of human control in the car should require not less connection and isolation but rather much more connection. In the absence of human intervention a driverless car should seek connection with navigation and advice systems such as weather systems, traffic infrastructure systems, signalling systems, other cars, central control systems, manufacturers' web sites and so on. Every type of connection should be pursued to compensate for the loss of human connection. I do not believe that the human social interaction and environmental awareness can be replaced by learning algorithms. Dialogue with the physical and human environment should be amplified not suppressed in a driverless car. The transfer of competency to the technology requires engagement with a wider knowledge base, not the exclusion of external information.

Driverless cars will. Of course, be subject to risks from security breaches, hacking and the compromising of privacy. However, the solution to the cybersecurity problem associated with autonomous vehicles is not to isolate oneself in total autonomy, to shut oneself down, but to open up and create strong communities of support, knowledge and cooperation to resist the threats.

5.3 Transparency

Transparency is a prerequisite for ethical engagement in the development of autonomous cars. There can be nothing hidden, no-cover-ups, no withholding of information. The limits of the driverless car, how it works and how it should be used should be made completely clear. Issues concerning safety, ethical decision making and the setting of boundaries cannot be addressed without transparency. There can be no deception, and no case of the robot car pretending to be what it isn't, creating an illusion of a capability it does not have. There is a difference between imitating a competence and actually having that competence.

The behaviour of Robotcar will depend on the learning algorithms. In the case of personal health monitors, the different algorithms used by the manufactures to turn electrical signals from sensors into data concerning number of steps and distance travelled can result is widely differing figures. And furthermore the meaning of those figures must then be determined. In another example, algorithms for turning the sequences of many short fragments of DNA into genome sequences can vary significantly in their results. The assemblathon competition [2] pits algorithms against one another to see which can come closest to giving the accepted sequence for a benchmark genome. A learning algorithm may varying in its learning and hence is response to environmental stimuli.

Hence clarity and openness about the algorithms used, how they work and their limitations is not only required technically, but must be communicated in a useful and appropriate way to users, managers, regulators and other interested parties. Limits in data localisation and interpretation need to be understood. The user of an autonomous car needs to be an intelligent user, knowing when to intervene, working with a human-centred interface.

Transparency may require a driving test for driverless cars to demonstrate their competency in navigation, dealing with roads and adhering to codes and laws in the particular geography and culture it will operate in. If we require testing of humans, we should require testing of robot cars.

5.4 Identity

In the cartoon film, Wall-E, the rule of technology renders humans passive, incompetent and hedonistic. Free of risk, responsibility and activity, they fail to engage with their environment, to question it and shape their future. They are willing, passive participants in an endless present. Obese, ignorant and unquestioning, they lounge by the swimming pool day after day. They surrender autonomy and responsibility to the technology, computer systems and robots. Additionally they had surrendered their identity. It is only in rebellion, catalysed by Wall-E, that they regain identity and purpose and return the technology to its rightful place. Taking on responsibility and embracing risk, they return to earth to start recolonisation.

Cars are often part of a person's identity. Not only the make and nature of the car, but the competencies in driving and the freedom and control the car provides, constitute part of the person's identity. The removal of competencies by the autonomous vehicle will clearly affect people's identity. In societies where reputation, wealth, and role in society are represented in the car and its use, driverless cars will pose a threat. Resistance to driverless cars may be partly driven by a fear of loss of identity. Social and personal identity may be undermined, or at the least transformed, by the driverless car.

Driverless cars may trigger identity crisis where the person is uncertain about his worth because his skills are transferred to the car. In a sense, by learning to imitate the driving skills and style of the human driver, the car is stealing part of the human's identity and becoming that person by imitating skills and roles which are part of who that person is.

Cars become part of people, extensions of them, gloves to fit into. The car fits the person, and is absorbed into the person. Stripping the person of such connection and involvement with the technology may not result in freedom but a fracturing of the person's identity which leaves them suspended in fear and uncertainty. Conversely for a disabled person, the driverless car may offer an extension of ability. The freedom which results from the ownership and use of a driverless car becomes an important part of that person's identity.

Will the autonomous car make people stupid? Will it steal a person's identity, taking on the driver's personality and characteristics? Understanding the role of the robotcar in human identity will required empathic reflection as well as an investigation of people's perceptions of the role of cars in their lives. People connect with cars. Controlling a car may be seen as a form of freedom. Loss of this human autonomy may equate to loss of identity.

5.5 Value

In discussing value we are interested in what people value. Value does not necessarily equate to benefits; it is not about cost benefit. Neither is it about values, our underlying moral drivers. Values will affect what we value. And an analysis of what we value will point to the values underneath. Freedom might be valued above safety, pleasure above health.

In the case of driverless cars, a concern will be on the value we put on the life, the needs and concerns of the users of the robotcar. Do we value external requirements of economics, of efficiency above internal value of promoting human flourishing and excellence? Do we value the driverless car as a statement of technological advance? Are we focussing on the system and the economics over quality of life, or promoting the market and the individual over community and cooperation?

There is a danger of devaluing the driver who becomes an object to be moved around, a set of inputs for the car learning to take over. In considering the ethics of driverless cars, we must also address the need to protect the privacy of information about how the driver drives a car and where and when.

5.6 Empathy

"If you don't believe this you need to leave .. this has to be a true thing." Paul Newman, Oxford University [5]

In contrast to impinging our view on the users of autonomous vehicles, an empathic stance requires that we view the deployment and use of the Robotcar through the eyes of the users. This requires us to cross the empathy gap, to put our feet in the shoes of the person sitting in the driverless car. A brief poll of family and friends will reveal a wide range of reactions to a driverless car. Some regard it with fear and revulsion. Wary enough of driven cars and the danger of the roads, the prospect of a driverless car is completely unacceptable. Others may view driverless as a novelty, and want to know 'how it works' out of interest or a need for assurance about the reliability of the technology. The latter point relates to a need for transparency and a reluctance to treat a driverless car as a black box initially and get into it without the sufficient knowledge as to its technology and its reliability.

For some males, the prospect of being driven around by a driverless car may bring about a primitive sense of emasculation.

However viewed, the driverless car elicit emotions and reactions which the engineer must be sensitive to. The engineer has to consider the fears and hopes of drivers; the way of thinking of drivers. There must be a respect for the human. Not treating them as a dangerous annoyance to be removed from the system.

Far from a disconnection with the human and the elimination of the human from the system, empathy requires an increased engagement with the human both in development and deployment. There must be a search for the human behind the driverless cars; a mindfulness of every person experiencing

the phenomenon, everyone connected with the products, services, infrastructure and usage. There must be a radical listening to the feelings and needs of the user which seeks not to impose a scientific absolute but to respond and adapt to the human heart, and to demonstrate responsiveness and adaptability. There must be an empathy with the perception of risk and vulnerability which new technology elicits. There must be a continuous conversation with the users which determines to reduce the empathy gap.

Questions such as the following should be a matter for reflection and investigation: What is the effect of loss of control, fear of car? What is the effect on the non-user? On those exposed to the car from other cars?

6. FINDING THE ETHICAL ROLE FOR DRIVERLESS CARS

Understanding the ethical context of a driverless car requires a systemic understanding of its place as part of the connectivity of transport and indeed society. Autonomy, as mentioned is not an ultimate goal. The driverless car supports and mediates relationships in the community. By enabling a journey for an elderly person, the car should enable connection. Rather than stripping away the autonomy of the driver, stealing his freedom and rendering him a passive recipient, the role of this technology should be seen as that of a support worker, compensating for some frailty of the human, where compensation is appropriate, enabling the human to use of car when physical or social constraints may have prohibited it.

The role of support worker does not eliminate risk and vulnerability. Rather it may create new risks and new dependencies. The value of the driverless car will be found not in environmental savings, efficiencies, the "saving of time", and the avoidance of accidents, but rather in the extent that it promotes relationality. It as Coeckelbergh suggests, the ultimate danger is non-relationality [3, p55] then the real ethical worth is in how the driverless car enables people to connect, strengthens communities, enables meetings and human interaction which might have been difficult or impossible before. The ethics of driverless cars is then an ethics of relationship and the impact of the driverless car on the human – human and human- machine relationship. And the key point in relationships is the interface, the boundary at which information is exchanged, understanding achieved, tasks agreed and roles carried out. Using driverless cars will be a matter of teamwork, of working together in the pursuit of common goals and purpose. The robotcar is a connected element in a connected universe, one element connecting to the whole of transport, working with rather than dismissing smart infrastructures, training, and so on. Contributing to the whole, reflecting our dependencies on each other.

7. CONCLUSION

Full autonomy is not only practically pointless, it is ethically pointless. The pursuit of such autonomy does not recognise the essential importance of interdependencies between humans and machines and that it is not a case of one or the other, but both. Indeed the splitting of human and machine, of what is perceived as uncertain and risky from the scientific, the assured, the provable, the separation of the rational and the emotional or even the material and the spiritual is a false dichotomy. The technological and the human are more entangled, impossible to prise apart and must be considered as a whole.

This requires a systemic view, which addresses systems and relationships, which recognises the place of driverless cars in a connected system, which is open to the study of complex relationships, both networked and hierarchical, which may give rise to emergent behaviour, and to physical, social and ethical issues which may be unexpected. Gaining an understanding of this will require the development of ethical dialogues between systems, communities and technology. Fundamentally, this requires a human-centred approach, a team approach, which examines the interdependencies between driverless cars and their users.

The pursuit of full robot autonomy is not a practical necessary nor a useful response to our needs and concerns; rather it is born of a philosophical view, underpinned by a particular perception of the human state.

8. REFERENCES

[1] Akerman, E. (2013) UK unveils affordable self-driving robot car. IEEE Spectrum. 19th Feb 2013. http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/uk-affordable-self-driving-robotcar/

[2] Bradnam, K.R. et al (2013) Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience, 2:10

[3] Coeckelbergh,M. (2013) Human Being @ Risk Springer Heidelberg.

[4] KPMG. 2014. Connected cars to deliver huge UK jobs boost, finds first UK study. http://www.kpmg.com/uk/en/issuesandinsights/articlespublications/newsreleases/pages/connected-cars-to-deliver-huge-uk-jobs-boost-finds-first-uk-study.aspx

[5] Intelligence Squared (2013) Smarter Mobility: An Evening of Debate. http://www.intelligencesquared.com/events/smarter-mobility-an-evening-of-debate/

[6] McBride,N ., "ACTIVE ethics: an information systems ethics for the Internet age, " Journal of Information, Communication and Ethics in Society, vol. 12, no. 1, 2014, pp. 21-43.

[7] Waldrop, M.M. 2015. Autonomous Vehicles: No Driver Required. *Nature.* 518 (2015), 20-23.

[8] Wilton, P. (2015) 8 Things about Oxford's driverless tech. http://www.ox.ac.uk/news/science-blog/8-things-about-oxford%E2%80%99s-driverless-tech

Cyber Education: towards a pedagogical and heuristic learning

Isabel Borges Alvarez
Universidade Autónoma de Lisboa
Rua Santa Marta 56
1169-023 Lisboa, Portugal
+351213177600
ialvarez@ual.pt

Nuno S. Alves Silva
Universidade Lusíada de Lisboa
Rua da Junqueira, 188 a 194
1349-001 Lisboa, Portugal
+351213611502
nsas@lis.ulusiada.pt

Luisa Sampaio Correia
ISCTE - Instituto Universitário
Av. das Forças Armadas
1649-026 Lisboa, Portugal
+351217903000
correia.mluisa@gmail.com

ABSTRACT
The constant and rapid investments in Information and Communication Technologies (ICTs) have allowed the growth in the quality of information response available within the internet which requires considering and addressing the physical, financial, socio-demographic, cognitive, design, institutional, political and cultural types of access. The main purpose of this paper is to revise actual and new emerging ICTs and the use of its application tools in Education which is dominated by the linear paradigm in interaction and information as interactivity is not being accepted as a guiding principle. The concept of e-learning rests on the idea that pedagogy technologically sustained includes enough knowledge with regard to the wishes of learning processes, which are a process of mind embedded in a culture and also challenges and not just concepts. Learning is a process that humans have been trying to master for many centuries. However, there are so many different ways to do the process that it is sometimes very hard to determine which one is the best of a given situation. One such type of learning is heuristic learning. Through this method the students should discover things for themselves, through problem solving, inductive reasoning, or simply by trial and error. Discovering things by yourself, knowing from experience rather than books. In many situations it seems that heuristic learning is the most suitable when one really believes in something when one experiences it himself. That is what heuristic learning is really all about. This type of learning model, in an online format, is tailored to the adult learner who typically has a sense of self-direction related to individual interests, goals, strengths, and previous experience. Also the pedagogical theory of connectivism was born as a response to very fast ICT development which strongly influences education and which approach to problem solving is based on the use of simulation animations, making students change parameters and verify or seek the problem solving, applying their need of intuitive searching by the heuristic method; information is assigned by image without the application of long texts. After a series of simulation experiments, the students verify their results with a calculation and a real experiment.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues: Ethics

General Terms
Human Factors

Keywords
Cyber education, e-learning, pedagogy, heuristics, ethics, equity

1. INTRODUCTION
"In the information society, citizenship can only be exercised through the positive involvement of citizens in systems and processes of information" and "incorrect or false information or its misuse can endanger people, their privacy and their freedom" [6]. Bearing this motto in mind it becomes important to ensure a set of citizenship's values and "social rights" such as the rights to education, to health and to justice. Education aims to personify "the beliefs, traditions, customs, rituals and sensibilities along with the knowledge of why these things must be maintained" [26] in [41]. In western countries one of the main political concerns is to ensure a better and broader education. However, economic and social models are constrains to the implementation of a successful and enlarged information society. Scandinavian countries standard policy for education is the student, taking into account that each one of them is an individual with specific needs to be satisfied to guarantee a solid growth of their capacities. Still it is expected that the principle of the universal education reduces poverty and increases knowledge. Anglo-Saxony countries policies go to the raise of the education that supports the demands of the XXI century [6]. It is accepted worldwide that e-learning is a precondition for future social and economic growth having knowledge as the key resource to e-learning [41]. Distributed knowledge must respect and obey to principles common to all humanity regarding social, cultural, ethical and individual (as the goal and motivation of each individual) issues. Hence it is understood that continuous learning is required. Once current knowledge is too vast for the time of each individual, the aim is rapid access to knowledge that is necessary 'here' and 'now' for a given problem and person. "Today, experts are people who know where to find information of immediate use and only the most up-to-date information is useful.... Knowledge has a half-life which gets shorter all the time." [22]. Also it is acknowledged that "The

economic and financial globalization, the globalization of technologies, as well as planetarium environmental issues should be accompanied by a moral and political conscience equal to what is at stake. It is precisely at times like these that it is needed philosophical and ethical frameworks able to help us to think about globalization and the failure of ideologies that we are witnessing [32]." New learning interactions that were not perceived before can now be facilitated such as the instantaneous access to global resources, the opportunity to publish to a world audience, to communicate with a diverse audience, and the ability to share and compare information, negotiate meaning and co-construct knowledge. Such activities emphasize learning as a function of interaction with others [9]. So, it is critical the achievement of an easily accessed network, with a well structured and feasible content for learning proposes. It will allow the learner to quickly address information that helps him to increase knowledge and achieve a heuristic value that drives him towards new ideas, new questions and new hypothesis. For that, ethical and cultur**al** issues must be carefully addressed.

2. CYBER EDUCATION

2.1 Why e-learning

"Most people are taught in groups; most learning is an individual experience. Learning is defined as what sticks; it is what remains years later" [45]. The traditional learning that has prevailed for decades within the higher education area has been gradually revolutionized by the computer-based learning regarding the opportunities in communication, interaction and collaboration [16] in [25]. Technology changes the way work is done, allows knowledge integration, and requires new skills and different levels of literacy. It promotes virtual organizations, knowledge share between peers; new forms of training are designed and delivered. It affects the way knowledge is managed and how organizations learn. The limitations vary in different circumstances depending on students and instructors need, who and where they are [29]. Today's life style has speeded up our needs for knowledge and competitiveness. The global share of knowledge, the research on advanced technologies such as nanotechnologies, the extensive curriculums to be learned by students, the accessibility to information and communication technologies (ICTs) forced traditional teaching to evolve and made virtual learning critical to overcome the dependency on space and time, allowing individual and collaborative learning experiences. Pressures related to the increasing changing demographics of the student population increase diverse cohorts, age, educational backgrounds, cultures and native languages [29]. E-learning is one of the terms used to refer to open and distributed learning activities as well as online learning (OLL), web-based training (WBT), Web-Based Learning (WBL), Distributed Learning (DL), Mobile Learning (m-learning) and so on. According to the American Society for Trainers and Developments (ASTD) e-learning has increasingly come to mean "Web-Enabled material deployed using the Net". E-learning can be delivered synchronously "live" in a virtual class-room in real-time (used for distance learning) or asynchronously by the download of contents made available by a tutor, any place at the convenience of the learner (used for distributed learning). **I**n other words "e-learning is nothing more than the use of electronic tools and technologies to assist us in our teaching and learning." [27], in [8].

2.2 Meaningful Learning

"Meaningful learning occurs when learners are able to remember, recall, understand and reuse the knowledge to explain to others or apply it in their everyday life" [47]. The associated theories related to pedagogy [47] are: Developmental theory; Learning theories; Cultural diversity; Classroom motivation and management; Learning styles; Instructional design; and Assessment.

Where some of these theories are briefly presented:

Developmental theory - This theory gives the foundation for teachers and instructors to understand their learners through the cognitive development ([30] describes the way people organize information and how the process changes during a child's development stages), through socio-cultural development ([44] emphasis the social and cultural influences on children's cognitive development) and finally the moral development [30] which provided the children's stages response to moral problems).

Learning theories - These theories come from the psychological theory and are used to understand critical issues rose in the learning process, such as mechanisms of learning and transfer, the roles of memory and motivation.

Learning Styles - It was described by Kolb as the individual's preferred method for assimilating information; it is related with the active learning cycle. It indicates how different styles affect the learner's performance. As learning styles can provide context to learning objects [47] it is of importance to refer some of the Models of Learning Styles. These are: Models based on the learning process; Models based on orientation to study; and Models Based on instructional preference.

2.3 Learning Methodologies

Teaching-learning strategies should incorporate more than one form of study. Methodologies such as Web-Quest involve the study of individual pre-defined web-sites followed by student's additional information research activities and review to guarantee that what he has learned becomes a part of his knowledge. This process allows learning share between work groups resulting in the construction of knowledge [35]. This type of process can occur synchronously via teleconferencing, chat or asynchronously through individual work that must be shared later. However, any process must be developed beyond a satisfactory e-learning experience. It is not enough to provide an efficient and effective e-learning environment, it needs to empower and motivate students to learn [23]. It should so focus on cognitive development and knowledge acquisition, through creative, efficient and intelligent tutoring strategies for presentation of the domain knowledge [1]. There are several evaluation methods to design a useful e-learning system. Michael Giannakos [13] focused on an e-learning system based on two cores: usability (ensuring usability is one of the main challenges for the e-learning developers) and pedagogical usability (which is divided into learning effectiveness and learning efficiency). Collaborative learning in contrast to traditional, lectured-base learning is "an interactive, group knowledge-building process" [16] in [25]. It instructs peers, not necessarily expertise in the same area, learners, teachers, researchers, professionals - to work together on a consensus among members of that community. They have a common goal and are opened for the share of their skill, experiences and knowledge. Collaborative learning within a real-world environment and within a virtual environment (such as an online collaborative learning) differs in size, place and method or form. A real-world environment takes place in a small group of participants composed by heterogeneous skills and abilities, face to face interaction takes place and they learn from more experienced colleagues or confluent areas of

knowledge. Rules for the teamwork are introduced. They work together or individually on a regular meeting base for leadership, team management, problem solving and conflict resolution [10], in [25]. Online collaborative learning can be described as information communication network. It is accessible to the group allowing interaction between learners **to** take place no matter distance, place and time. Multicultural people meet in these networks for exchange of ideas and experiences and to accomplish their tasks within a common propose. Today's face-to-face interactions are available through virtual conferences rooms and the support of many management tools. "Collaborative learning environments (ENVITE, C-VISions) and 3-D environments (CLEV-R) which allow for asynchronous and group learning have been used over recent years" [28]. According to [12] synchronous tools supporting voice communication (Skype) can be considered a critical factor in enhancing group collaboration because voice adds a personal touch to the communication process [25]**.** Multidisciplinary skills and innovative forms of learning are often required. It must be adjusted to one's needs and reality. There is the need to rebind knowledge so it is important to identify the common characteristics of complex systems. As said by Joel de Rosnay [36], it is no longer only at the microscope or the telescope dimension, but also at the macroscope dimension. Both analytical and systemic approaches are complementary of one another. Within the analytical approach we hold the perception of detail, regardless of the time variable. Within the systemic approach we cover up an integrated view. While in the analytical approach only one variable is modified a**s** the time and facts are validated within an experimental theory, in the systemic approach groups of variables are modified simultaneously and the facts are determined by comparing a model with reality. Here computers, with its unlimited storage capacity and high speed data processing, play a role rich in the study of the humanities and technology.

2.4 Pedagogical Perspectives

According to Silva et al. [38], learning is an active process that aims to connect learner's new and old knowledge, mainly if it is an independent and lifelong learning that can involve three main formal perspectives: pedagogy, andragogy and heutagogy. Traditional pedagogy is the art of being a teacher where teacher-centered philosophies of education are essential. On the other hand, contemporary pedagogy is becoming more complex since it explores the methodologies enabled by the use of ICT in education. The inherent characteristics of interaction and flexibility in the networked environment impose new frontiers to whether individual or social learning occurs in formal and informal ways. In addition, artificial-intelligence evolution and new tutoring systems are as closely intertwined to explore virtual pedagogical models, which may include the relationship between culture and online learning behaviours. However, the application of learner-centred learning techniques is ensuring the flexibility and relevance of the curriculum personalization remaining underdeveloped, or at least constantly critical in the timeline of fast technological developments. In this sense, andragogy explores how to motivate adults to lifelong learning, but it is not clear in what extent digital literacy are imposing limits to that motivation or change in adult educational objectives. Conversely, heutagogy is a form of self-determined learning with practices and principles rooted in andragogy. In a heutagogical approach learners are highly autonomous and should develop learning path to be well-prepared for the complexities of today's society [5]. In that sense, Hase & Kenyon [17] noticed that heutagogy applies a holistic approach to developing learner capabilities, with learning as an active and proactive process, and learners serving as "the major agent in their own learning, which occurs as a result of personal experiences" (p. 112). In spite of personal experiences, the affordances provided by emerging technologies and the ubiquitousness of Web are renewing the interest in heutagogy, but the absence of critical culture to the creation and development of knowledge is an essential question. As Hase [17] noted, an important characteristic of heutagogy is that of reflective practice, or "a critical learning skill associated with knowing how to learn" (p. 49). Then, heutagogy is advocating principles of information literacy, critical thinking, and collaboration. Furthermore, each individual learner has different goals and characteristics from each other, leading to different associations between the learning content and the knowledge level of each learner, or a different learning path that represent a sequence of concepts and activities that the learner chooses or must be chosen during the learning process [20]. Consequently, personalized curriculum is an emerging phenomenon that can be supported by heutagogy and a concept map of heuristic information constructed in a network environment. So, it is critical the achievement of an easily accessed network, with a well structured and feasible content for learning purposes. It will enable the learner to quickly address information that helps him to increase knowledge and achieve a heuristic value that drives him towards new horizons. For that, ethical and cultural issues must be carefully addressed.

2.5 Heuristic Learning

Following Levy and Razin [24], the heuristic is based on the assumption that individuals share a common priority, transmit the whole distribution of their beliefs to one another and update solely on this information. The sharing principles dominate education in cyberspace, as well as, interactivity is a major objective in the use of modern technologies. However, learners should discover things by themselves, since intuitive searching is applied by the heuristic method. Anderson [4] also recalls that the heuristic learning model, presented in an online format, fits instruction tailored to the adult learner who typically has a sense of self-direction related to individual interests, goals, strengths, and previous experience. Conversely, the degree of uncertainty in a heuristic learning is relative to the individual expertise and mental models. Moreover, emerging technologies in educational settings are imposing the same process on young and old. While it will be good for cultural adaptation, it sounds a disaster for equity of knowledge, if learners value an item of information differently depending on who else knows it [15]. Moving forward issues of cyber education towards a pedagogical and heuristic learning means to put in analysis of the learners' autonomy versus the regulatory purposes of accredited education. A question emerges: How this type of learning can be made available to the wider global population to achieve equity of knowledge and progress?

3. E-LEARNING CONCEPTS

The key concepts in an e-learning project are: lecturer, content, student, place, time and interactivity. To guarantee the effective inputs for effective e-learning the following inputs must be taken into consideration [3]:

Visual - for instance relevant image give support to a simple text;

Concise - concise information is important in the context of e-learning;

Interactive - learners interact with multimedia activities during the courseware;

Engaging - appeals to all intervenient (learner's professional experience). Their emotional reaction may lead them to motivation;

Relevant - addresses learner's gaps or current needs;

Feasibility - technological infrastructure must be feasible to learners;

Empowering - provides access to additional resources as relevant material to explore.

Technological e-learning systems or subsystems may be classified into four categories [3]:

Learning Management Systems (LMS) - supports administrative tasks such as registration, scheduling and learner tracking;

Managed Learning Environment (MLE) - includes the whole range of information systems and processes, contributes directly or indirectly to learning and learning management;

Learning Content Management Systems (LCMS) - allow developers to store and manage and provide access to pieces of content used in e-learning;

Virtual Learning Environments (VLE) - The components in which learners and tutors participate in several online interactions, including on-line learning.

Further, the American Society for Trainers and Development (ASTD) refers to Learning Management System (LMS) as the "Operating System" for e-learning enterprises and it won't be wrong to consider LMS as the Operating System in any context where learning issues are involved and they are assessed through systems information built for that [19]. There are dozens of companies offering served-based LMS as pure repositories providers like Click2Learn and others as content providers. Some LMS are enhanced with Content Management Systems (CMS) functionality. All CMS or LCMS (Learning Content Management Systems) are developing compliance with the content object standards, such as SCORM (Sharable Content Reference Model). These combined efforts enable:

- learning objects easily reused;

- accessibility to learning objects developed by any proprietary software;

- portability and roll out facilitated;

- granular learner assessment models.

4. E-LEARNING FRAMEWORKS

E-learning concerns with successful projects ranging from:

- product quality (fitness for intended learning, appropriate design, intuitive navigation and fast, appropriate technology, response speed);

- availability and palatability of the learner (ease of handling systems from the standpoint of the user, curiosity, capacity sharing, and innovation);

- publicizing the correct courses by promoting organizations, adequacy of exercises timings, clarity of language, possibility of sharing knowledge.

To give support to the design and implementation of e-learning projects some ICT researchers have proposed their own frameworks. These frameworks' goal is to provide integrated guidance for design, development, delivery and evaluation of an e-learning environment. Some of these are briefly presented below. They are: RIPPLES Model, E_University framework and Khan's framework.

4.1 RIPPLES Model

RIPPLES model drawn by Dan Surry [42] is another implementation framework for embedding innovative practice in e-learning within the Vocational and Technical Education (VTE) sector in Australia where the monogram stands for:

Resources - the need for continuing resources, temporary resources and resources allocation;

Infrastructure - the hardware, software, facilities and network capabilities;

People - shared decision-making and communication between all stakeholders;

Policies - organizational policies and procedures to adapt to innovative practice;

Learning - the need for innovation to enhance the training goals of the organization;

Evaluation - the need for continual assessment of the innovations;

Support - the need to have a support system in place for those implementing the innovation.

4.2 E-University Framework

e-University framework engages an interactive real time feedback process on e-learning implementation in higher education settings [39]. It is characterized by four progress layers as described:

Technological Infrastructural and Services (support and services) - encompass all technical support and administrative services for the distributed knowledge.

Knowledge/Content Management (production and distribution) - it emphasizes content and knowledge production, management and distribution through multiple technological platforms;

Computer Mediated Communication (interactive technologies) - it makes learning more interactive and an enjoyable experience;

Value Added - a transversal cost/benefit analysis, which aims to provide information regarding the e-University project at a final stage.

The strength of this model is the constant evaluation of equity by cost/benefit of each layer and its related cultural and ethical issues if the project involves multiple contexts.

4.3 Khan's Framework

Khan [21] has come forward with his own model, specific to e-learning environment. It focuses on eight dimensions, as presented below:

Pedagogical (teaching and learning) - this dimension addresses analysis for: content, audience, media, goal, design, organization, methods and strategies of e-learning environment;

Technological (infrastructure) - this dimension concerns are systems availability, interoperability and maintenance costs;

Interface design - this dimension concerns are page and site design, navigation, content design and usability. The quality of a page design is based on how user-friendly, appellative, easy to read the site is;

Evaluation - this dimension includes assessment of learners and instruction and learning environment;

Management - this dimension is related to the maintenance of learning environment and distribution of information. A good site must be currently updated and constantly reviewed by experts on the subject;

Resource support - this dimension is related to on-line support and resources required to foster learning's environments;

Ethical - this dimension takes into account considerations related to social and political influence, culture diversity, bias, geographical diversity, learner is diversity, information accessibility and legal issues;

Institutional - this dimension is related to administrative affairs, academic affairs and students' services.

Mohammed Ally [2] in his work for best practice and standards indicates Khan's Framework as an important tool. He also advises to develop e-learning material as learning objects so it can be accessed from any computer technology and can be reused in different lessons or courses.

5. CONTENT: A TRADE-OFF ANALYSIS

When we see human inability to address the current amount of information as well the theories that support a given knowledge, the creation of content is needed and must follow [3]**:**

- a consensus view on existing theories for each topic;

- a content that must be sufficiently explicit;

- it should facilitate the creation of new theses (heuristic value);

- a satisfactory 'marriage' between knowledge, experience and characteristics of learners.

Indeed while modernism brought us expertise needed to encompass knowledge, postmodernism favours the encounter of knowledge in a consensual and entire form. Knowledge has paths that intersect and from where cognition emerges [37].

5.1 Content Certification

When looking at contents it is important to ensure the quality and certification of its sources' and guarantee that its contents are regularly reviewed and updated. For instance, we may be accessing harmful and manipulative information without knowing it. This kind of information may reach us in an apparently harmless way. Easily when playing poker games with "virtual" partners in the net they frequently use compliments to promote the potential addition of the individual to the game, driving families to serious bankruptcy. Also, some concerns still subsist related to the content creation and maintenance. Who is to certify the quality of the information within these databases? What standards will define the massive creation of information and its updating? What are the criteria to determine the relevance of information? Who has the author's rights, the person who created it or the enterprise for which she works? What determines the maturity of a learner to confine in him his own learning? Are learners grown up enough to behave ethically, to understand the barriers of privacy, culture issues? How to ensure the equity of knowledge of all who access these data? The dialogue between local and global ethical systems (glocal ethics) suggest a mutual and equal respect, thus higher education institutions have a social responsibility to promote "glocal knowledge" and so a concomitant recognition of "glocal morality" [40].

5.2 Tools and Infrastructures

These e-learning distributed platforms allow users to create and manage classified information made available for at least two groups: students who access courses and teachers who are responsible for the creation and updating of the course's structure, its contents, evaluation system. Also, e-learning platforms allow students and teachers 'discussion boards', 'chat exchange', e-mail, instant-messaging, video-conferencing, monitoring training, questionnaires and interactive exercises as well as to assess reports and surveys for evaluating actions. Administrative tools for management and for assessing are available for user management and content management, allowing the creation and edition courses. Students and teachers enrolment are also available. Still, "e-learning offers one–location gateway to varieties of educational resources, such as electronic book, digital presentation, web-based lecture notes, case studies and other types of educational learning materials" [47]. These digital materials need to be built from scratch using past experience for guidance. The portability of systems has been possible through the communication technology. The Learning Objects (LO) concept introduces small, portable learning materials on the Internet. The utility of LO is the reusability of the objects in practically any environment. The repository where these objects are is a Learning Object Metadata (LOM). IEEE LTSC (2005) is one of the standards used as the benchmark in LO metadata Development [47].

Today's most common tools for face-to-face or distance learning are Moodle (acronym for Modular Object-Oriented Dynamic Learning Environment), Power-Point and many games once integrated in a multimedia platform which allows intercommunication between users. Moodle is an open source software adopted by a majority of colleges and universities as a way to communicate and share information with their students either in a classroom or remotely, such as http://www.schoolanywhere.co.uk/. Power Point learned in schools has become a common tool used by teachers, and students to present their work. It can integrate voice, documents, text and movement. For multimedia platforms there are products such as Adobe Flash and Dreamweaver. Dreamweaver is currently applied to the use of web-pages design, with hypertext links allowing navigation between information as it suits the user. On other hand, Flash is highly used in the creation of design of web pages and movies. At present**,** games and simulators are being developed. School curriculum units applied to these games and simulators are aimed in the engagement of student to the topic, fastening his cognitive growth. The Global Challenge World Game is intended to provide pre-college students the opportunities for self-instruct: science, technology, engineering and mathematics. The World Game uses the Microsoft ESP visual simulation platform to turn available to students "immersive 3-D

experience" designed with the propose of helping them to understand complex nature of global systems. Each curriculum will be inspired in game experience simulation. The Microsoft ESP was chosen due to its fast simulation construction and effective cost. A digital game and simulation-based approach to STEM (acronym for the fields of Science, Technology, Engineering and Mathematics) learning both accommodate student preferences and support the core cognitive process of learning [14].

5.3 Future Tendencies on E-learning

The heading of subsections should be in Ti xxx Traditional classroom course based training will remain and be shared with technology-based learning, mobile learning using laptops, tablets, PDAs and cell phones. Often learning takes "forgetting" and "relearning" new ways of thinking and doing things. Computer simulation will master real situations. Games are an ideal way to introduce people to new topics. It engages people into play and learning. Dramatic changes in technology such as the constant growth of capacity and velocity allows larger networks computers. As molecular computer evolves and nanotechnologies and methods of dividing light into specific wavelength communication channels proliferate. "The introduction to artificial intelligence and neural networking will make e-learning software smarter and more responsive. New online learning programs will be both prescriptive and adaptative." [45]. It will allow the computer to learn more about its user, his needs and preferences; it will 'sense' his behaviour and will provide him with his learning and testing needs. Like computers, in any type of machine such as cars, traffic control is invisible, computer training will also be invisible. Small devices hooked to a network will perform tasks and perform learning activities, such as household devices. It will also deliver human resources information (health information), business metrics and so on. All this will be driven by artificial intelligence [45] such as:

Automatic computing - computers will self-control its resources by configuring, healing and so on;

Agent-based software - web search engines used for planning, notifying and negotiating;

Affective computing - computer software will sense emotions and act accordingly. "They will increase the realism of e-learning simulations." [45].

The FH JOANNEUM University department of information design [11] created a prototype AdeLE (Adaptative e_Learning with Eye tracking) resulting from its past experience in hypermedia and application of eye tracking for web usability evaluation in the Web Usability Center. The eye tracking is applied for more profound learning research and improvement of cognitive processes understanding to be able to support adaptive teaching and learning in a technology-based e-learning in the future [31]. Data from (i) learning, (ii) reading (iii) skimming through text, (iv) searching in the text, (v) observing a picture or reading a text and (vi) looking on the navigational elements are reported to the prototype simultaneously with real-time eye tracking. It assesses the learning state and it enhances a user profile for learning style of user, cognitive style: holist or analyst. The entire content is presented in different ways: holist style an overview of chapters and sub-chapters are optionally offered while for the analyst style the whole content is presented. The first issue of these methods is to extract individual learning strategies. People exhibit significant individual differences in how they learn [31]. AdeLE framework can be integrated into different applications such as content management systems in e-learning environments.

6. SOCIAL AND ETHICAL IMPACTS

Our thanks to ACM SIGCHI for allowing us to modify templates they had developed.xxx Spending time on the internet is changing our behavior and culture which is referred as cyber-culture. "The Cyber-culture is not simply a culture of cyber-space and navigation in the huge resources of information; it is a culture of global government…. What is new here is that cyber-culture uses means of our time to act on problems of our time [32]. Information technology is nowadays the most prominent technological development that affects our everyday life, and our dependence on information technology increases constantly. As a consequence, emerging ethical issues that individuals or professionals face, require appropriate skills [33]. Cyber-ethics is a new terminology to refer to ethical concerns with property rights, privacy and correct use and divulgation of information in the cyber world. This subject is referred as cyber-education. Due to its capital importance it is a subject to be considered as part of the education and learning process; it should involve teachers and students from early years of school. Computer illiteracy of staff and students are factors that compromise the correct use of ICTs as an educational technology. According to Philipe Quéau [32] culture is "what can give each person reasons to live and to wait" It's what gives new means to increase the beauty and wisdom of the world ... culture ... lives of breaths, streams, fertilization and miscegenation ... ". It is this understanding of education and culture that enforces the need of man for constant learning and improvement. No doubt, it is a global and local obligation of governors and citizens to take measures to:

- make the access to ICTs as broad as possible;

- the effective effort towards the use of ICTs must be done in the same way that yesterday was made in relation to reading;

Take into consideration the positive and negative aspects of the current state of ICT and Education and assure they have their own ethical tools regardless the censorship they are subjected to (although you can do a clone of Man, it is not due to censorship reasons that it is not done but due to consensual reasons).

6.1 Equity

To guarantee more equitable global society cyber-education should be enhanced to balance cultural and ethical issues and anticipate problems to come. Moreover, the evolution of cyber or all related future technologies have the potential to change cultures and ethical questions may arise [3]. Thus, the constant evaluation of equity at a level of the project progress assumes a major importance if it involves worldwide contexts, since cultural and ethical differences have relevant impacts. The local and global ethical systems interplay the learning process at networks that cannot be separated from the knowledge creation [39]. Therefore, it is important to distinguish the heuristic learning from pedagogies based on certified content. Or, at what extend the purpose of education depends on the ways that knowledge recipient is available to collaborate in social processes that enhance excellence and equity of knowledge. For example, while ICT is equal at a global scale, concerning education certain values build large communities that overpass national boundaries but only make regional impacts (e.g. African ubuntu). On the other hand, there are generation gaps that should be considered in terms

of equal competence and digital literacy, which may be the source for inter-generational inequity [43]. In these scenarios, cultural and ethical impacts include the way how cyber-education is distributed equitably across learners, and if the learners' life-chances are enhanced in equitable ways.

7. DISCUSSION AND CONCLUSION

The rate at which technology diffusion takes place is astonishing and is altering how and to whom we are connected. The wireless connected machines within the internet turn these repositories into endless metadata. People need to learn much faster and are in a constant learning process and consequently seeking for information. Unquestionably, e-learning will continue to grow in our organizations and at schools and universities. Remote teaching may be an advantage for those who live far from colleges and want to improve their knowledge. Governments and enterprises must work on effective and efficient solutions. Feedback management is a key process to the success of any e-learning platform. Teachers must be responsible for his students' first steps at the net. They must teach the young to be responsible for their actions and to share information with others. It is an ethical principle to guarantee others safety, privacy, equity and equality when it comes to workspace. The freedom of each must halt the moment you get freedom of others and vice versa [32]. Many barriers are still to overcome (interactivity, procurement practices, policies, performance) so it may be used fully by students at colleges or universities and enterprises. Many questions are still unanswered. The virtual 3D environment and neural computers are not far from now. It will revolutionize the way e-learning will be done. It will be possible to manipulate objects, to interact with the computer. It will learn from us and optimize its own its processes. We will be able to perform and to teach remotely surgery. These kinds of interventions will be of great precision due to the diagnosis at hand and the precision of computer assistance. For all these future technologies and changes of culture we must center on ethical behavior questions. Mainly we must guarantee equity and equality and walk as much as possible towards a balanced information society and try to anticipate problems to come. Cognition results from the combination of the integrity and the experience of knowledge [37].

8. REFERENCES

[1] Ahmad, A., Basir, O., and Hassanein, K. 2004. Adaptive User Interfaces for Intelligent E-learning: Issues and Trends. In *Proceedings of the The Fourth International Conference on Electronic Business* (ICEB 2004). Beijing.

[2] Ally, M. 2011. Best Practices and Standards for e-Learning. Paper presented at *2nd International Conference on e-Learning and Distance Learning*. Riyadh, Saudi Arabia. Available at http://eli.elc.edu.sa/2011/sites/default/files/slides/%20%D8%A2%D9%84%D9%8A.pdf

[3] Silva, N., Costa, G., Prior, M., and Rogerson, S. 2013. The Evolution of E-learning ManagementSystems: An Ethical Approach. In K. Beycioglu (Ed.), *Ethical Technology Use, Policy, and Reactions in Educational Settings* (pp. 93-106). Hershey, PA: Information Science Reference. DOI=10.4018/978-1-4666-1882-4.ch008

[4] Anderson, S. 2008. Effective Integration of Sound Pedagogy in an Online Format. In J. Luca & E. Weippl (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology 2008* (pp. 3579-3586). Association for the Advancement of Computing in Education (AACE). Available at http://www.editlib.org/p/28882.

[5] Blaschke, L. M. 2012. Heutagogy and Lifelong Learning: A Review of Heutagogical Practice and Self-Determined Learning. *The International Review of Research in Open and Distributed Learning*, 13, 1.

[6] Coelho, J. D. 2011. *Do Plano Tecnológico à Agenda Digital: Cinco anos de tomadas de posição do grupo de alto nível da APDSI*. Edições Sílabo.

[7] Costa, G., and Silva, N. 2010. Knowledge versus content in e-learning: A philosophical discussion! *Information Systems Frontiers*, 12, 4, 399-413.

[8] Costa, G., Silva, N., and Fonseca, T. 2012. Moral reasoning in knowledge authoring: an e-learning 4.0 analysis! In S. Abramovich (Ed.), *Computers in Education- Volume 1* (pp. 135-154). Hauppauge, NY: Nova Science Publishers.

[9] Dabbagh, N. 2005. Pedagogical models for e-learning: a theory based design framework, *International Journal of Technology in Teaching and Learning*, 1, 1, 25-44

[10] Felder, R., Woods, D., Stice, J., and Rugarcia, A. 2000. The Future of Engineering Education, Part 2. Teaching Methods that Work. *Chemical Engineering Education* 34, 1, 26–29.

[11] FH JOANNEUM 2014. *AdeLE (Adaptive e-Learning with Eye-Tracking): Theoretical Background, System Architecture and Application Scenarios*. FH JOANNEUM University, Austria. Available at http://researchanddesign.fh-joanneum.at/wp-content/uploads/2014/03/adele_folder_en_0.pdf

[12] Franceschi, K., Lee, R., and Hinds, D. 2000. Engaging E-Learning in Virtual Worlds: Supporting Group Collaboration. In: R. H., Sprague, Jr. (Ed.), *Proceedings of the 41st Hawaii International Conference on System Sciences*, Waikoloa, Big Island: IEEE Computer Society Press.

[13] Giannakos, M. 2010. The Evaluation of an E-learning Web-based Platform. In *Proceedings of the 2nd International Conference on Computer Supported Education*. CSEDU '10. INSTICC Press, 433-438.

[14] Gibson, D., and Grasso, A. 2008. *An enterprise simulation platform for education: Building a world game for pre-college students with Microsoft ESP*. Microsoft white paper. Available at http://www.microsoft.com/education/highered/whitepapers/simulation/simulationplatform.aspx#games

[15] Glazer, R. 1998. Measuring the knower: Towards a theory of knowledge equity. *California Management Review*, 40, 3, 75-94.

[16] Harasim, L., Hiltz, S. R., Teles, L., and Turoff, M. 1995. *Learning Networks, A Field Guide to Teaching and Learning Online*. Cambridge MA. The MIT Press

[17] Hase, S., and Kenyon, C. 2007. Heutagogy: A child of complexity theory. *Complicity: An International Journal of Complexity and Education,* 4, 1, 111-119.

[18] Hase, S. 2009. Heutagogy and e-learning in the workplace: Some challenges and opportunities. *Impact: Journal of Applied Research in Workplace E-learning,* 1, 1, 43-52

[19] IsoDynamic 2001. e-Learning Whitepaper. Available at http://www.isodynamic.com/web/e_learn.htm

[20] Kardan, A. A., Ebrahim, M, A., and Imani, A. M. 2014. A new Personalized Learning Path Generation Method: Aco-Map. Indian J.Sci.Res. 5, 1, 17-24.

[21] Khan, B. 2001. *A framework for e_learning*. Available at http://www.elearningmag.com/elearning/article/articleDetail.jsp?id=5163

[22] Kearsley, G. 2000. *New Developments in Learning*. Available at http://home.sprynet.com/~gkearsley

[23] Lera, E., and Mor, E. 2007. *The joy of e-learning: redesigning the e-learning experience*. Project PERSONAL (TIN 2006-15107-COZ-01), Barcelona

[24] Levy, G., and Razin, R. 2014. *A simple Bayesian heuristic for social learning and groupthink*. London School of Economics and Political Science. Working paper. Available at http://personal.lse.ac.uk/levyg1/learning.pdf

[25] Lukman, R., and Krajnc, M. 2012. Exploring Non-traditional Learning Methods in Virtual and Real-world Environments. International Forum of Educational & Society (IFETS). *Educational Technology & Society*, 15, 1, 237-247.

[26] Maison, K. B. 2007. A case for professional studies in education for teachers in higher educational institutions. In D.E.K. Amenumey (Ed.), *Challenges of education in Ghana in the 21st century* (pp. 248-256). Accra: Woeli Publishing Services.

[27] Martin, E., and Webb, D. 2002. Is E-learning good learning? *The ethics and equity of e-learning in higher education* (pp. 49-60). Melbourne: Victoria University: Equity and Social Justice.

[28] Monahan, T., McArdle, G., and Bertolotto, M. 2008. Virtual reality for collaborative e learning. *Computers and Education*, 50, 1339–1353.

[29] Omwenga, E. I; Waema, T., and Wagacha, P. W. 2004. A model for introducing and implementing e-learning for delivery of educational content within the African context, *African Journal of Science and Technology*, 5, 1, 34-46.

[30] Piaget, J. 1964. Cognitive development in children: The Piaget papers. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget rediscovered: A report of the conference on cognitive studies and curriculum development* (pp. 6-48). New York: Cornell

[31] Pivec, M., Trummer C., and Pripfl, J. 2005. Eye-Tracking Adaptable e_learning and Content Authoring Support. University of Applied Sciences FH JOANNEUM, Graz, Austria. *Informatica*, 30, 83-86.

[32] Quéau, P. 1999. *O Desafio do Século XXI – Religar os Conhecimentos: Cybercultura e Info-etica. Estomologia e Sociedade.* Instituto Piaget.

[33] Rigopoulos, G., and Karadimas, N. V. 2006. Increasing Ethical Awareness of IT Students through Online Learning. In *Proceedings of the 6th WSEAS International Conference on Applied Informatics and Communications,* 265-269.

[34] RIPPLES Model 2015. Available at http://designplanet.wikispaces.com/RIPPLES+Model

[35] Romiszowski, A. 2003. *O futuro de e-learning como inovação educacional: fatores influenciando o sucesso ou o fracasso de projetos*. Associação Brasileira de Educação a Distância. Revista Brasileira de Aprendizagem Aberta e a Distância, S.Paulo, Novembro, 2003.

[36] Rosnay, J. 1999. *O Desafio do Século XXI – Religar os Conhecimentos: Conceitos e Operadores Transversais*. Instituto Piaget.

[37] Santos, B. S. 1988. *Um Discurso sobre a Ciência* (A Discourse on the Sciences). Porto Afrontamento.

[38] Silva, N., Costa, G., Rogerson, S., and Prior, M. 2009. Knowledge or content? The philosophical boundaries in e-learning pedagogical theories! In *Proceedings of the m-ICTE 2009* (pp. 221-225). Lisbon. Portugal

[39] Silva, N., Alvarez, I., and Rogerson, S. 2010. *Glocality, diversity and ethics in distributed knowledge to Higher Education*. In G. Costa (Ed.), *Handbook of Ethical and Social Issues in Knowledge Management: Organizational Innovation* (pp. 131-159). Hershey: IGI Global.

[40] Silva, N., Rogerson, S., and Stahl, B. C. 2010. Ethicultural sensitivity in e-learning: discussing Lusíada Universities empirical findings. In *Proceedings of the ETHICOMP 2010* (pp. 500-511). Tarragona. Spain.

[41] Silva, N., Costa, G., Prior, M., and Rogerson, S. 2011. The evolution of E-learning Management Systems- an ethical approach. *International Journal of Cyber Ethics in Education*, 1, 3, 12-24.

[42] Surry, D. W., Ensminger, D. C., and Haab, M. 2005. A model for integrating instructional technology into higher education. *British Journal of Educational Technology*, 36, 2, 327–329. DOI=10.1111/j.1467-8535.2005.00461.x

[43] der Have, V., and Nienke, S. 2013. *The Right to Development and State Responsibility can States be Held to Account?* Symposium held in honor of the tenth anniversary of the Prince Claus Chair in Development and Equity. Amsterdam Law School Research Paper No. 2013-23;. Available at SSRN: http://ssrn.com/abstract=2251838

[44] Vygotsky, L. S. 1978. *Mind in society*. Cambridge: Harvard University Press.

[45] Woodill, G. 2004. *Seven trends in corporate eLearning*. White paper. Operitel Corporation. Available at www.operitel.com.

[46] Yahya, Y., and Yusoff, M. 2005. The perception of a learning object model, its characteristics and metadata: From theoretical perspectives. In *Proceedings of the E-Learn Conference. Vancouver, Canada.* 24 – 28 October 2005.

[47] Yahya, Y. and Yusoff, M. 2008. Towards a comprehensive learning object metadata: Incorporation of context to stipulate meaningful learning and enhance learning object reusability. *Interdisciplinary Journal of E-Learning and Learning Objects*, 4, 13-48. Available at http://www.ijello.org/Volume4/IJELLOv4p013-048Yahya185.pdf

Digital wildfires: hyper-connectivity, havoc and a global ethos to govern social media

Helena Webb
Marina Jirotka
University of Oxford, Department of Computer Science
Oxford, United Kingdom
helena.webb@cs.ox.ac.uk
marina.jirotka.cs.ox.ac.uk

Bernd Carsten Stahl
De Montfort University, Department of Informatics
Leicester, United Kingdom
bstahl@dmu.ac.uk

William Housley
Adam Edwards
Matthew Williams
Cardiff University, School of Social Sciences
Cardiff, United Kingdom
housleyw@cardiff.ac.uk
edwardsa2@cardiff.ac.uk
williamsm7@cardiff.ac.uk

Rob Procter
University of Warwick, Department of Computer Science
Coventry, United Kingdom
rob.procter@warwick.ac.uk

Omer Rana
Pete Burnap
Cardiff University, School of Computer Science and Informatics
Cardiff, United Kingdom
ranaof@cardiff.ac.uk
p.burnap@cs.cardiff.ac.uk

ABSTRACT
The last 5-10 years have seen a massive rise in the popularity of social media platforms such as Twitter, Facebook, Tumblr etc. These platforms enable users to post and share their own content instantly, meaning that material can be seen by multiple others in a short period of time. The growing use of social media has been accompanied by concerns that these platforms enable the rapid and global spread of harmful content. A report by the World Economic Forum puts forward the global risk factor of 'digital wildfires' – social media events in which provocative content spreads rapidly and broadly, causing significant harm. This provocative content may take the form of rumour, hate speech or inflammatory messages etc. and the harms caused may affect individuals, groups, organisations or populations. In this paper we draw on the World Economic Forum report to ask a central question: does the risk of digital wildfires necessitate new forms of social media governance? We discuss the results of a scoping exercise that examined this central question. Focusing on the UK context, we present short case studies of digital wildfire scenarios and describe four key mechanisms that currently govern social media content. As these mechanisms tend to be retrospective and individual in focus, it is possible that further governance practices could be introduced to deal with the propagation of content proactively and as a form of collective behaviour. However ethical concerns arise over any restrictions to freedom of speech brought about by further governance. Empirical investigation of social media practices and perspectives is needed before it is possible to determine whether new governance practices are necessary or ethically justifiable.

Categories and Subject Descriptors
K.4 [**Computers and Society**]: Public and Policy Issues – *abuse and crime involving computers, ethics, regulation.*

General Terms
Management, Human Factors, Legal Aspects.

Keywords
Social media, governance, responsible research and innovation.

1. INTRODUCTION
The last 5-10 years have seen a massive rise in the popularity and use of social media platforms such as Twitter, Facebook, Instagram, Snapchat and Tumblr etc. A 2014 report prepared by the UK's independent regulator OFcom [1] found that over 80% of British adults go online regularly and that 66% of these have a social media profile. Social media platforms enable users to post their own content – news, opinions, images etc. – which is then available to be seen instantly either publicly or by audiences selected by the user. Most of these platforms also have functions that allow users to forward some of the content they see, through shares or retweets etc. This content therefore has the capacity to

be seen by multiple others across the globe in a short period of time.

This rapid spread of content via social media can offer undoubted societal benefits, such as the promotion of social cohesion through solidarity messages and humanitarian campaigns [2]. However as social media platforms have grown significantly in popularity, concerns have also risen over their capacity to enable the rapid spread of harmful content. Reports of 'cyber-bullying', harassment and 'shaming' on social media have become commonplace in popular media [3], whilst governments and other institutions have blamed platforms such as Twitter and Facebook for enabling the spread of false rumours [4] and inciting violence [5] during times of tension. These concerns have led to calls for more effective regulation of digital social spaces [6] – for instance through the criminalisation or restriction of certain kinds of user content. Inevitably however these calls are contradicted by other arguments that position the internet as a medium that supports and encourages freedom of speech and therefore label any increased regulation as unethical [7].

In this paper we take up these contemporary concerns over the propagation of content on social media and the appropriate governance of digital social spaces. We draw on a 2013 report by the World Economic Forum (WEF) [8], which describes the global risk factor of 'digital wildfires' – social media events in which provocative content of some kind spreads broadly and rapidly, causing significant harm. We discuss the WEF's report further in Section 2 and highlight a central question arising from it: does the risk of digital wildfires necessitate new forms of social media governance? In Section 3 we present the findings of a scoping exercise conducted to begin addressing this central question. Focusing on the UK context we present three short Case Studies of digital wildfire scenarios and then characterise the four key governance mechanisms relevant to the regulation of these scenarios. We identify gaps within current governance and in Section 4 suggest potential further practices that might be adopted to overcome them. We highlight ethical issues surrounding the adoption and justification of any new governance mechanisms. We also argue that empirical research is necessary to analyse the real time propagation of content on social media and investigate the practical experiences and perspectives of various stakeholders in the governance of digital social spaces. This empirical work will be taken up by the paper authors in further project work.

2. BACKGROUND

2.1 Social media and digital wildfires

In February 2013 the World Economic Forum published the report "Digital wildfires in a hyperconnected world" [see 8]. As part of the Global Risks series, the report describes the popular use of social media platforms as a serious threat to international security and societal well-being. Social media platforms enable information and misinformation to spread rapidly and reach huge audiences, so where this content is in some way provocative – for instance taking the form of rumour or hate speech, or containing politically or socially inflammatory messages – it can 'wreak havoc in the real world'. The report conceptualises these risks as 'digital wildfires': social media events in which provocative content of some kind spreads broadly and rapidly, causing significant harm.

The WEF report gives examples of potential digital wildfire scenarios. It describes how the spread of misinformation can cause harm because it has negative consequences before there is an opportunity to correct it. For instance the spread of unverified content can damage the reputation of an individual – as in the false naming in 2012 of a senior UK politician in connection to allegations of child abuse (see Case Study 1). It can also undermine the standing of commercial companies, organisations, or institutions - such as in false reports of British Army failures in Iraq in 2009. Furthermore it can undermine social cohesion, for instance by causing panic over apparent incidents of disease outbreaks and security threats or by reinforcing the 'groupthink' of individuals who position themselves in networks separate from the rest of society.

2.2 Digital wildfires and the governance of social media

The WEF report describes digital wildfires as arising from the 'misuse of an open and easily accessible system'. Social media platforms are widely and freely available to many users across the world and place relatively few constraints against provocative content in the form of an unverified rumour, inflammatory message etc. Given the negative consequences that this spread of provocative content can cause, the report asks whether digital wildfires can be prevented through effective governance. It notes that legal restrictions on freedom of speech are technically difficult to achieve internationally and ethically difficult to justify. Instead it argues that as digital social spaces continue to evolve, there is scope for the development of a 'global digital ethos' in which generators and consumers of social media adopt *responsible practices*. The development and promotion of this ethos are challenges that remain to be undertaken.

2.3 New practices towards a global ethos to govern social media?

The World Economic Forum's description of digital wildfires provides a useful means to conceptualise the risks posed by the propagation of provocative content on social media. Digital wildfires can be understood as fast paced phenomena involving a form of collective behaviour through the spread of content by multiple users. They can result in significant harms and present challenges to the effective and ethical governance of digital social spaces. If we accept digital wildfires as a global risk factor, we are led to examine the role of governance in regulating the 'havoc' they can cause and the potential for a global ethos promoting digital responsibility. Therefore the WEF report prompts a central question: does the risk of digital wildfires necessitate new forms of social media governance?

The remainder of this paper reports on a scoping exercise designed to begin answering this central question. Focusing on the UK context, we reviewed current social media governance relevant to digital wildfire scenarios. Through a series of case studies and the examination of relevant literature and resources, we identified four key governance mechanisms: legal governance, social media governance, institutional governance and user self-governance. We then identified the characteristics of these mechanisms and highlighted a number of gaps in their capacity to deal with digital wildfire scenarios. Whilst it may be possible to introduce further governance practices to fill these gaps, our scoping exercise reveals the need for further empirical investigation to determine whether new mechanisms are in fact necessary or ethically justifiable.

3. SCOPING THE CURRENT GOVERNANCE OF SOCIAL MEDIA

3.1 Rationale and questions for scoping exercise

The scoping exercise was conducted as part of an ongoing research project on the responsible governance of social media (see Section 5.3.2). We drew on the World Economic Forum report to pose a central question: does the risk of digital wildfires necessitate new forms of social media governance? In order to address this question we determined that it was also necessary to consider further questions:

- What governance mechanisms currently exist relevant to digital wildfires?
- How do current governance mechanisms map on to potential digital wildfire scenarios?
- Are there any gaps in current governance mechanisms?
- Could any gaps in current governance be filled by new mechanisms?
- (How) can new governance mechanisms be ethically justified?

3.2 Conduct of scoping exercise

Governance practices, in particular legal frameworks, can vary across countries. In order to produce specific findings that could map directly on to particular digital wildfire scenarios, we decided to focus on social media governance in the UK (where the project is funded and based). We identified a number of social media events that could be categorised as potential digital wildfire scenarios. We conducted case studies of these scenarios to identify: the kind of provocative content propagated across social media; the governance mechanisms applied and their impact; and questions and debates arising over the appropriate regulation of the scenario. Three of the case studies are summarised in Section 4.2.

Through the case studies we identified four key mechanisms that seem to operate in relation to digital wildfire scenarios in the UK: legal governance, social media governance, institutional governance, and user self-governance. We examined each mechanism in turn through reference to news reports, institutional reports and reviews, websites and social media platform Terms of Use etc. We assessed the scope of these existing mechanisms and identified gaps in their capacity to prevent or manage digital wildfire scenarios. We then identified a range of further governance practices that could potentially overcome these gaps. As these further practices might be seen to limit freedom of speech, this then led us to highlight important ethical considerations surrounding the regulation of digital social spaces. Finally, we reflected on our findings in relation to the central question posed by the scoping exercise.

4. THE CURRENT GOVERNANCE OF SOCIAL MEDIA IN RELATION TO DIGITAL WILDFIRE SCENARIOS

4.1 Overview of findings

In this section we present the results of our scoping exercise and describe the current governance of social media in relation to potential digital wildfire scenarios. We begin with short summaries of three digital wildfire case studies. We then identify and discuss the characteristics of four key governance mechanisms: legal governance, social media governance, institutional governance, and user self-governance.

Our results indicate that legal governance, social media governance and institutional governance all tend to be retrospective in character; they deal with the kinds of provocative content associated with digital wildfires after it has spread and had an impact. They also tend to act on individual users rather than the multiple users who may be involved in a digital wildfire. By contrast user self-governance appears to have a real time element and may have the capacity to limit the spread of content posted by individuals or multiple users.

4.2 Case studies of digital wildfire scenarios

In the first stage of the scoping exercise we identified events meeting the criteria of digital wildfires: that is, they involved the rapid and broad spread of some kind of provocative content on social media which caused significant harm to an individual, group, organisation and/or population. We drew up case studies of these scenarios to identify the different mechanisms that were applied to regulate the digital wildfire.

Three of the case studies are summarised here. They have been chosen as they exemplify: 1) the kinds of content that may be involved in a digital wildfire; 2) the different kinds of governance mechanisms that may be drawn on to regulate a digital wildfire; and 3) current debates around the appropriate regulation of digital social spaces.

4.2.1 Case Study 1: Lord McAlpine

On 2nd November 2012 a BBC television programme broadcast a report on the sexual abuse of children in North Wales care homes during the 1990s [9]. It revealed that two of the care home victims had identified a "leading politician from the Thatcher years" as one of their abusers. The broadcast did not name the politician concerned but – alongside subsequent reports from other news media – provided enough information to enable many people to infer that it referred to Lord Alistair McAlpine. People began to name him on social media - including Sally Bercow, political activist and media personality with over 55 000 followers. She posted the tweet shown in Box 1.

In the week following the broadcast it became apparent that McAlpine had been wrongly implicated in the report [10]. The BBC issued an apology and subsequently paid McAlpine £185 000 in damages. Some Twitter users immediately issued apologies for naming him. McAlpine and his legal team considered reporting the Twitter messages naming him to the police and then announced they would sue users for libel [11]. Experts were hired to collate all relevant tweets: around 10 000 tweets were identified as potentially defamatory – 1 000 original tweets and 9,000 retweets.

Box 1. Tweet posted by Sally Bercow

Ultimately, users with fewer than 500 followers were asked to make a charitable donation in return for having cases against them dropped and McAlpine announced his attention to pursue libel actions against 'high profile' users with more than 500 followers [12]. Whilst out of court settlements were reached with a number of these high profile figures, Bercow maintained that her tweet was not defamatory and the case was taken to court. At the trial, Bercow's argument that her tweet constituted a 'random' thought was rejected and the judge found that her reference to 'innocent face' was insincere and ironical [13]. The case was formally settled in October 2013. Bercow apologised for her 'irresponsible use of Twitter' and agreed to pay McAlpine undisclosed damages and cover his costs. She then temporarily closed her Twitter account. The case attracted a great deal of attention in the UK and was referred to by McAlpine's lawyer as "the leading case in terms of internet responsibility" [14].

4.2.2 Case Study 2: Caroline Criado-Perez

Caroline Criado-Perez is a journalist and feminist activist who was involved in a successful and high profile campaign in spring 2013 to guarantee a place for female historical figures (in addition to Queen Elizabeth II) on banknotes produced by the Bank of England [15]. Following the campaign, Criado-Perez wrote an article in the *New Statesman* revealing that she had been receiving numerous rape threats via Twitter from multiple accounts [16]. She reproduced some of the content of the tweets in the article (without including the account handles of the users who sent them) – see Box 2. Criado-Perez reported the tweets to the police and strongly criticised Twitter for not doing enough to deal with the threatening messages and the users who posted them.

"this Perez one just needs a good smashing up the arse and she'll be fine"

"*Everyone jump on the rape train > @CCriadoPerez is conductor"; "Ain't no brakes where we're going"*

"Wouldn't mind tying this bitch to my stove. Hey sweetheart, give me a shout when you're ready to be put in your place"

Box 2: Examples of abusive tweets quoted by Caroline Criado-Perez

The article provoked a range of discussion over the appropriate ways to deal with online harassment [17]. Some argued that reporting abuse to the police or social media platforms was unnecessary as users could 'use their own voices' to shame others who harassed them. However Criado-Perez maintained that the police and Twitter needed to do far more to help victims of harassment. A petition started in July 2013 calling for Twitter to simplify and speed up its systems for reporting abuse received 40 000 signatures in its first week [18]. In August 2013 the head of Twitter UK apologised to Criado-Perez for the abuse she had received and pledged that the platform would do more to stop similar abuse occurring [19]. Twitter subsequently introduced a 'report tweet' function that enabled users to report abuse immediately rather than having to send a message through its Help Centre [20].

In January 2014 Isabella Sorley and John Nimmo pleaded guilty to sending menacing messages to Criado-Perez [21]. It was stated in court that Criado-Perez had received abusive messages from 86 Twitter accounts, including multiple accounts held by the two defendants. It was also reported that Criado-Perez had suffered life changing psychological effects from the abuse she had received. Both Sorley and Nimmo received prison sentences and were described by their defence lawyers as naïve in their use of social media, taken in by the attention they received when their abusive posts were retweeted, and unaware of the harms they had caused.

4.2.3 Case Study 3:2011 England riots

On August 6th 2011 a peaceful protest over the police shooting of a man in south London became violent [22]. Over the next few nights disorder and looting spread across London and other towns and cities in England. Social media platforms such as Twitter and Facebook were widely used during this period and were seen by the government and some other commentators to play a significant role in enabling the spread of rumour, incitement of violence and organisation of gang activity.

The riots resulted in over 3 000 criminal prosecutions and a number of these involved the use of social media. For instance, Perry Sutcliffe-Keenan [23] received a 4 year custodial sentence after pleading guilty to intentionally encouraging another to assist the commission of an indictable offence. On August 9th Sutcliffe-Keenan had used his Facebook account to invite his 400 followers to riot in the town of Warrington the following day. However, he deleted the page shortly after setting it up and subsequently described it as a joke. No riots occurred in the town but the page was reported to the police by some members of the public. The court was told that Sutcliffe-Keenan's actions had caused panic in the local area and placed a strain on police resources. In another example a 17 year-old youth [24] was banned from social media sites for 12 months and ordered to complete 120 hours of community service after admitting sending a menacing message that encouraged rioting. He had posted a Facebook message saying "I think we should start rioting, it's about time we stopped the authorities pushing us about and ruining this country. It's about time we stood up for ourselves for once. So come on rioters – get some. LOL." The court heard that some of the youth's followers who saw the message replied by calling him an 'idiot' for posting it and the youth had deleted it by the time the police arrived to talk to him about it. No riots took place in the area where the youth lived and he told the court that the post had been intended as a joke.

The England riots prompted a great deal of discussion about the impact social media messages can have on offline behaviours and how/whether this should be governed. On August 11th Prime

Minister David Cameron announced that the government would review the possibility of preventing suspected rioters sending messages online [25]. In response to criticism of the site, a Facebook spokeswoman confirmed that the platform removed 'credible threats of violence' as part of its monitoring process [26]. She also pointed to the positive role that Facebook played during this time of great tension by providing a means for users to let family and friends know they were safe. Subsequent research [27] has suggested that the impact of social media in escalating the riots was overestimated; BBM smart phone messaging was used to coordinate illegal activity far more than social media and the response of Twitter and Facebook users to the unfolding events was more anti than the pro the riots. Many individuals took to social media to send messages condemning the violence and used the platforms to coordinate 'clean up' operations after the riots had ended.

4.3 Key governance mechanisms relevant to digital wildfires

The collation of case studies of digital wildfires enabled us to identify four key governance mechanisms relevant to digital wildfires. The characteristics of these governance mechanisms are discussed in turn.

4.3.1 Legal governance

In July 2014 the UK House of Lords Select Committee on Communications published a review of Social Media and Criminal Offences [28]. This concluded that, with the exception of criminalising online behaviours associated with 'revenge porn', it was not necessary to introduce new laws to govern social media in England and Wales. Therefore, legal actions regarding social media draw on existing civil and criminal legal codes. These can pursue individuals who have posted certain kinds of provocative content – such as defamatory claims (Case Study 1), menacing or obscene messages (Case Study 2) incitements to violence (Case Study 3), threats of violence, and breaches of court orders. Punishments for breaking these laws take the form of fines/damages, community service and custodial sentences.

In a typical digital wildfire scenario, a relatively small number of potentially illegal posts are reported to the police/lawyers and an even smaller number of these are pursued in the courts. In Case Study 1 the vast majority of users reached out of court settlements with the lawyers representing Lord McAlpine. In Case Study 2 the police were unable to identify all the users who had posted menacing messages and some cases were dropped as pursuing them was deemed not to be in the public interest [29]. In Case Study 3 only a very small number of users who posted inflammatory content about the riots were reported to the police.

Legal actions deal with provocative social media content retrospectively, after it has been posted, spread and had an impact. Beyond the use of deterrent sentences, legal governance therefore has little capacity to prevent the spread of provocative content and digital wildfires. Rhetoric around legal governance has frequently emphasised the limitations of the law in dealing with mass postings on social media [30]. It has also emphasised the responsibility of individual users to behave appropriately on social media (Case Study 1) and understand the potential impacts of their actions (Case Study 2).

4.3.2 Social media governance

Although social media platforms differ in the precise ways that they govern user content and behaviour, social media governance typically centres on the application of Terms of Use agreements. Platforms such as Twitter, Facebook, Flickr, Instagram, Tumblr etc. require users to sign up for an account by providing some contact and/or identifying information and agreeing to follow specific Terms of Use regarding what they can and cannot post on the platform. The Terms of Use generally set out penalties for breaches in the form of deletion of posts and suspension or closure of accounts.

Automated processes can identify and block certain types of content, such as explicit threats of violence (Case Study 3) and images of child sexual exploitation. However most often platforms rely on other users to report breaches of the Terms of Use. In some cases social media companies may pass on information to the police or security services, although they can be reluctant to do so [31].

Social media platforms often promote user self-governance. In addition to being able to report others, platforms typically have privacy and blocking functions so that users can control who has access to their posts. Certain features on a platform can encourage trust amongst users. For instance the use of real names and/or the addition of demographic information can help users to feel they 'know' each other. Users may also have the option to rate, rank, 'like' or 'favourite' others' posts to indicate that they – and by extension the user that posted them – are creditworthy. Similarly, users can draw on information about how many friends, followers etc. a poster has or how many posts they have made to draw conclusions about that poster's trustworthiness. Finally, some of the large social media service providers have taken part in awareness and education campaigns to promote responsible user behaviour [32].

The governance mechanisms of social media platforms are still evolving and changes are made on a regular basis. Twitter brought in significant changes to its reporting process following the abuse of Criado-Perez (Case Study 2) and has introduced further steps to tackle 'trolls' in 2015 following criticism from its own CEO [33]. However Twitter, like other social media platforms, is underpinned by the principle of freedom of speech and explicitly states that it upholds the right for users to post inflammatory content [34]. Sally Bercow's tweet in Case Study 1, although defamatory, did not breach Twitter's Terms of Use and the posts in Case Study 3 were not treated (at that time) by Facebook as credible threats of violence.

As with legal governance, the governance mechanisms within social media platforms focus on dealing with individual users and posts. Therefore they lack capacity to deal with the multiple posters involved in a digital wildfire scenario. Automated processes can to prevent the posting and reposting of certain kinds of content but most breaches are dealt with retrospectively and rely on user reports. As reporting can be a slow process, provocative posts can be often be seen and shared repeatedly – potentially causing significant harm - for a considerable period before they are acted on.

4.3.3 Institutional governance

As social media sites have grown in prevalence and popularity, organisations of various kinds have begun to institute policies to

govern appropriate content and user behaviour relevant to the particular institution. For example various employers require their employees to follow policies that outline what can and cannot be posted in official and personal accounts [35]. Typically, these place constraints on the posting of (negative) information about the employer organisation and can also extend to penalising users who undermine the organisation by behaving inappropriately – for instance by posting racist comments. Guidance to jurors in the UK now incorporates the use of social media [36] and many schools set out social media protocols to be followed by staff, students and parents [37]. Institutional governance appears to have some capacity to deal with the kinds of provocative content associated with digital wildfires as social media policies are likely to sanction certain kinds of unverified and inflammatory content. But once again this form of governance tends to be retrospective in focus and acts on individual users and posts after content has been spread.

4.3.4 Social media user self-governance

Users can undertake a number of actions that function to govern social media content. Where applicable they can report posts to the police or social media platform (Case Studies 2 and 3) or pursue other users through civil law (Case Study 1). They can set up privacy settings etc. to monitor who has access to their posts. They can delete or alter their own posts (Case Study 3) and even suspend their accounts (Case Study 1) where appropriate.

Users can also challenge content posted by others. For instance they might label a post as misleading or inappropriate. In Case Study 1 some of Bercow's followers urged her to remove her defamatory tweet and apologise for it before the trial, whilst work conducted on the 2011 riots found that users were able to successfully challenge and limit the spread of unverified rumours [38]. An alternative kind of challenge is to mock the poster in order to minimise the value of a post. For instance in Case Study 3 some followers labelled the youth an 'idiot'. Taken further, users also sometimes seek to 'shame' users for posting inappropriate content. This can be done in a variety of ways and includes: encouraging others to criticise a user; finding and spreading identifying details of the user; and passing on the user's posts to monitoring sites such as 'Yes, you're racist' or 'Racists getting fired'. Shaming can be highly effective in the sense that it can lead to users leaving the social media platform or losing their job etc. but it does raise ethical concerns over whether the harm it inflicts is justified by the harm caused in the offending post [39].

Finally, ignoring provocative posts and users has long been advocated as a way to deal with inappropriate content [40]. It stops content being spread and deprives users of the attention they are seen to crave. However since many social media posts have a very wide reach, it is perhaps unlikely that a large number of users will all ignore a provocative post. Furthermore some victims of online harassment (Case Study 2) argue that it is important to fight back against provocative posts rather than letting them pass without comment.

Self-governance practices appear to have some prospective characteristics. They may be able to counter the provocative content associated with digital wildfires in real time – for instance by challenging and correcting misinformation or preventing the spread of posts. Exactly how these practices play out during digital wildfires is a question that requires empirical investigation.

5. DISCUSSION

In this section we discuss the implications of the results of our scoping work. We describe gaps in current governance relating to digital wildfires and suggest further governance mechanisms that may overcome these gaps. We highlight key ethical questions regarding the introduction of any further governance practices and conclude that more empirical research is necessary to address our central question – does the risk of digital wildfires necessitate new forms of social media governance?

5.1 Current governance related to digital wildfires

5.1.1 Characteristics

We identified four current governance mechanisms relevant to digital wildfires: legal governance, social media governance, institutional governance and user self-governance. These mechanisms differ in the kinds of content they treat as inappropriate and in the kinds of sanctions they apply but all map on to digital wildfire scenarios to some extent.

Legal, social media and institutional governance mechanisms tend to be retrospective in focus as they deal with content after it has been posted. They typically apply sanctions to individual users. By contrast self-governance mechanisms have a real time element and may limit or prevent the spread of some posts.

Rhetoric surrounding these various mechanisms shares an emphasis on the importance of responsible user behaviour and can be seen to reflect the interest of the World Economic Forum in the development of a digital ethos that moves beyond legal regulation.

5.1.2 Gaps in current governance

None of the four governance mechanisms deal with digital wildfires as a specific phenomenon so it is inevitable that gaps in current governance arise. A key gap concerns the capacity for governance practices to act on multiple users rather than individuals. As described by the World Economic Forum, digital wildfires can be understood as involving a form of collective behaviour through the cumulative spread of content by multiple users. Legal, social media and institutional governance procedures focus on individual users and/or posts and therefore lack the capacity to deal with this characteristic. In addition, as these mechanisms – apart from the use of automated processes by social media platforms to block some kinds of content – deal with content retrospectively they do not have the capacity to prevent or limit the impact of digital wildfires in real time.

5.2 Potential further governance mechanisms

5.2.1 Types of mechanisms

It is possible that further governance structures could be introduced to map more directly onto the characteristics of digital wildfires and overcome some of the gaps noted above. This could include:

- Technical mechanisms to counteract the rapid spread of social media content. For instance the creation of a waiting time for retweets that could be linked to activity around a post or user. This would be comparable in principle to measures that slow down automatic trading when markets behave erratically.

- Further support for self-governance mechanisms that challenge and slow down the spread of provocative content. For instance the provision of visible esteem to individuals who intervened in the early stages of a digital wildfire in order to ensure the appropriate spread of content. Alternatively, the provision of a 'lie' button to indicate that the content of a post is not creditworthy or an 'ignore' button that users can activate to recommend that others do not respond to a post.
- Automated content analysis of posts to identify potentially defamatory, misleading, offensive etc. content. This could then trigger a warning to users recommending review of the post before submission.

5.2.2 Justification of governance

The alternative governance mechanisms suggested above are designed to limit the development and spread of digital wildfires and reduce the impact they can have. This is based on the assumption that digital wildfires are harmful and need to be limited. Insights from computer ethics and responsible research and innovation [41] illustrate the importance of ethical justifications for governance and in the case of digital wildfires this is not straightforward. Questions of harm and truth are central. Preventing the spread of provocative content can be beneficial but some mechanisms may produce more harm than the content itself. For instance, preventing or delaying the posting of content could be seen as a significant barrier to freedom of speech – and this in turn, as the World Economic Forum report acknowledges, can have very negative consequences. In addition the increasing prevalence of social media 'shaming' of posters can appear out of proportion to the harm done in an offending post. In any case how can the truthfulness or potential harmfulness of a post be established – and by whom? Wildfires that are based on truthful content may well be desirable – even if the content is provocative in other ways. Any consideration of governance mechanisms that limit digital wildfires needs to balance considerations of freedom of speech with issues concerning the avoidance of harm. This is in part a normative question but is also one that can be informed by empirical insights into how provocative content spreads on social media, the harms it causes and the capacity for existing governance mechanisms to deal with it.

5.3 Further questions

5.3.1 Need for empirical research

The results of our scoping exercise highlight the existence and characteristics of four key governance mechanisms operating in the UK context. We have shown that gaps in governance exist and that further governance practices may be possible but that these require careful ethical examination.

However this scoping exercise alone cannot answer the central question regarding the regulation of digital social spaces in the context of digital wildfires. Further questions emerge from our work which require empirical investigation. How do existing governance mechanisms operate in real time in digital wildfire scenarios? In particular, what role does self-governance play in limiting and halting the spread of provocative content? Furthermore, what kinds of harm do digital wildfires inflict on different individuals, groups, organisations and populations? Are these harms serious enough to support arguments for new mechanisms that will potentially limit freedom of speech? A better empirical understanding of digital wildfires is required to determine whether new governance mechanisms are necessary and justified, and what forms they might take. This important empirical work is taken up by the authors in our ongoing project – "Digital wildfire: (Mis)information flows, propagation and responsible governance."

5.3.2 The "Digital Wildfire" project

The "Digital Wildfire: (Mis)information flows, propagation and responsible governance" project [42] is an interdisciplinary study led by the University of Oxford in collaboration with the Universities of Cardiff, de Montfort and Warwick. The overall aim of the project is to build an empirically grounded methodology for the study and advancement of the responsible governance of social media in the context of digital wildfires. The scoping work discussed in this paper forms part of a review of existing governance mechanisms which will inform the empirical activities of the study. The empirical work will take 3 forms: 1) Case studies of 4 digital wildfires. We will collect digital media datasets and combine computational analysis with qualitative analysis to examine information flows during digital wildfires and the occurrence of self-governing behaviour, such as counter speech to combat rumour and antagonistic content. 2) We will conduct a series of online questionnaires to seek the informed opinion of various experts regarding the appropriate regulation of digital social media and digital wildfires. 3) We will conduct interviews and observations at various sites (such as social media platforms, police organisations, civil rights groups) to investigate and understand how stakeholders respond to instances where the digital spread of provocative content may create situations of offline tension, conflict or disturbance.

The results of the scoping and empirical work will be drawn on to produce an ethical security map. This will be a practical tool to help different users navigate through social media policy and aid decision making. Other project outputs include the development a training module on digital maturity and resilience for use in secondary schools and the production of artwork to promote a creative understanding of digital wildfires amongst a broad range of audiences.

6. CONCLUSION

In this paper we have drawn on the concept of digital wildfires – social media events in which provocative content spreads broadly and rapidly, causing significant harm – and reported on a scoping exercise conducted to investigate a central question: does the risk of digital wildfires necessitate new forms of social media governance? We have described and discussed existing governance mechanisms relevant to digital wildfires in the UK context and identified a number of gaps in current governance. We have highlighted opportunities for further governance practices that could overcome these gaps by prospectively preventing and limiting the spread of provocative content. We have also highlighted ethical concerns around the introduction of any new governance practices that might limit freedom of speech. The question of whether new governance approaches are necessary to regulate digital wildfires requires further investigation; we have demonstrated the need for empirical research that analyses the real time propagation of provocative content on social media and investigates practical issues and perspectives regarding its governance.

7. ACKNOWLEDGMENTS

The "Digital Wildfire: (Mis)information flows, propagation and responsible governance" project is funded by the Economic and Social Research Council. Project reference ES/L013398/1.

8. REFERENCES

[1] Ofcom. 2014 *Internet Citizens Report 2014* (Nov. 2014), DOI= http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/Internet_Citizens_Report_14.pdf

[2] Rotman, R., Vieweg, S., Yardi, S., Chi, E., Preece, J., Shneiderman, B., Pirolli, P. and Glaisyer. T. 2011. From slacktivism to activism: participatory culture in the age of social media. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '11). ACM, New York, NY, USA, 819-822. DOI=http://doi.acm.org/10.1145/1979742.197954.

[3] Ronson. J. 2015. *So You've Been Publicly Shamed*. (March 2015) Riverhead Books, London. 277 pages.

[4] Luckerson, V. 2014. Fear, misinformation and social media complicate ebola flight. t*ime.com* (Oct. 2014). DOI= http://time.com/3479254/ebola-social-media/

[5] Halliday, J. (2011) David Cameron considers banning suspected rioters from social media. *theguardian.com* (Oct. 2011) DOI= http://www.theguardian.com/media/2011/aug/11/david-cameron-rioters-social-media

[6] All party parliamentary group against anti-Semitism. 2015. *Report of the all party parliamentary inquiry into anti-Semitism* (Feb. 2015). DOI= http://www.antisemitism.org.uk/wp-content/themes/PCAA/images/4189_PCAA_Antisemitism%20Report_spreads_v9%20REPRO-DPS_FOR%20WEB_v3.pdf

[7] Foxton, W. (2015) Criminalising online trolls is absurd even if what they say is vile. t*elegraph.co.uk* (Feb. 2015) DOI= http://www.telegraph.co.uk/technology/social-media/11401308/Criminalising-online-trolls-is-absurd-even-if-what-they-say-is-vile.html

[8] World Economic Forum, 2013 *Digital Wildfires in a hyperconnected world*. Global Risks Report. World Economic Forum (Feb. 2013), DOI= http://reports.weforum.org/global-risks-2013/risk-case-1/digital-wildfires-in-a-hyperconnected-world/

[9] Greenslade, R. 2014. Newsnight's McAlpine scandal. 13 days that brought down the BBC's chief. *theguardian.com* (Feb. 2014). DOI= http://www.theguardian.com/media/greenslade/2014/feb/19/newsnight-lord-mcalpine

[10] Sabbagh, D. and Deans, J. 2014. BBC to pay Lord McAlpine £185 000 damages after false child abuse allegations. *theguardian.com* (Nov. 2012) DOI= http://www.theguardian.com/politics/2012/nov/15/bbc-lord-mcalpine-compensation-newsnight

[11] Swinford, S. 2012. Lord McAlpine vows to take on the "Twittering fraternity" *telegraph.co.uk* (Nov. 2012) DOI= http://www.telegraph.co.uk/culture/tvandradio/bbc/9688845/Lord-McAlpine-vows-to-take-on-Twittering-fraternity.html

[12] Dowell, B, 2012. McAlpine libel. 20 Tweeters including Sally Bercow pursued for damages. *theguardian.com* (Nov 2012). DOI= http://www.theguardian.com/tv-and-radio/2012/nov/23/mcalpine-libel-bercow-monbiot-davies

[13] Lord McAlpine of West Green v Sally Bercow. 2013. EWHC 1342 (QB). DOI= https://www.judiciary.gov.uk/judgments/mcalpine-bercow-judgment-24052013/

[14] Press Association. 2013. Lord McAlpine libel row with Sally Bercow formally settled in high court. *theguardian.com* (Oct. 2013) DOI= http://www.theguardian.com/uk-news/2013/oct/22/lord-mcalpine-libel-row-sally-bercow

[15] Bell, M. 2013. Victory for women on banknotes campaign. *Sofeminine.com* (July 2013) DOI= http://www.sofeminine.co.uk/key-debates/victory-for-women-on-banknotes-campaign-jane-austen-to-be-on-10-note-s86135.html

[16] Criado-Perez, C. 2013. After the Jane Austen announcement I suffered rape threats for 48 hours, but I'm still confident the trolls won't win. *New Statesman online* (July 2013). DOI= http://www.newstatesman.com/media/2013/07/after-jane-austen-announcement-i-suffered-rape-threats-48-hours-im-still-confident-tro

[17] Gold, T. 2013. How do we tackle online rape threats? *theguardian.com* (July 2013) DOI= http://www.theguardian.com/commentisfree/2013/jul/28/how-to-tackle-online-rape-threats

[18] Miller, B. 2013. UK petition calls on Twitter to tackle abuse after Caroline Criado-Perez subjected to violent tweets. *abc.net* (July 2013). DOI= http://www.abc.net.au/news/2013-07-29/thousands-sign-petition-to-stop-abusive-tweets/4849780

[19] BBC news. 2013. Twitter's Tony Wang offers apology to abuse victims. *bbc.co.uk* (Aug. 2013). DOI= http://www.bbc.co.uk/news/uk-23559605

[20] Doshi, S. 2014. Building a safer twitter. *Blog.twitter.com* (Dec. 2014). DOI= https://blog.twitter.com/2014/building-a-safer-twitter

[21] Cockerell, J. 2014. Twitter 'trolls' Isabella Sorley and John Nimmo jailed for abusing feminist campaigner Caroline Criado-Perez. *independent.org.uk* (Jan 2014). DOI= http://www.independent.co.uk/news/uk/crime/twitter-trolls-isabella-sorley-and-john-nimmo-jailed-for-abusing-feminist-campaigner-caroline-criadoperez-9083829.html

[22] Gentleman, A. 2011. London riots: social media helped gangs orchestrate looting, says MP. *theguardian.com* (Aug 2011). DOI= http://www.theguardian.com/uk/2011/aug/11/riots-social-media-gang-culture

[23] BBC News. 2011. England riots: two jailed for using Facebook to incite disorder. *bbc.co.uk* (Aug. 2011). DOI= http://www.bbc.co.uk/news/uk-england-manchester-14551582

[24] Miller. L. 2011. UK riots: 17 year old banned from using social networking sites for Facebook message. *Mirror.co.uk* (Aug. 2011). DOI= http://www.mirror.co.uk/news/technology-

science/technology/uk-riots-17-year-old-banned-from-social-185122

[25] Halliday, J. (2011) David Cameron considers banning suspected rioters from social media. *theguardian.com* (Oct. 2011) DOI= http://www.theguardian.com/media/2011/aug/11/david-cameron-rioters-social-media

[26] Halliday, J. and Garside, J. 2011. Rioting leads for Cameron to call for social media clampdown. *theguardian.com* (Aug. 2011). DOI= http://www.theguardian.com/uk/2011/aug/11/cameron-call-social-media-clampdown

[27] Lewis, P., Newburn, T., Taylor, M., Mcgillivary, C., Greenhill, A., Frayman, H. and Proctor, R. 2011. *Reading the Riots: investigating England's summer of disorder.* The London School of Economics and Political Science and The Guardian, London, UK. DOI= http://eprints.lse.ac.uk/46297/1/Reading%20the%20riots%28published%29.pdf

[28] House of Lords. 2014. *Social Media and Criminal Offences: 1st report of Session 2014-2015*. London. The Stationery Office Limited. DOI=: http://www.publications.parliament.uk/pa/ld201415/ldselect/ldcomuni/37/3702.htm

[29] Smith, J. 2013. Two to be charged with threatening tweets to campaigner who called for a woman to be on bank notes. *Mailonline* (Dec. 2013). DOI= http://www.dailymail.co.uk/news/article-2524784/Two-charged-threatening-tweets-woman-banknotes-campaigner-Carolina-Criado-Perez.html

[30] House of Lords. 2014. *Social Media and Criminal Offences: 1st report of Session 2014-2015*. London. The Stationery Office Limited. DOI=: http://www.publications.parliament.uk/pa/ld201415/ldselect/ldcomuni/37/3702.htm

[31] Independent Reviewer of terrorism legislation. 2015. *A question of trust. Report of the legislator powers review*. June 2015. DOI= https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/434399/IPR-Report-Web-Accessible1.pdf

[32] House of Lords. 2014. *Social Media and Criminal Offences: 1st report of Session 2014-2015*. London. The Stationery Office Limited. DOI=: http://www.publications.parliament.uk/pa/ld201415/ldselect/ldcomuni/37/3702.htm

[33] Tiku, N. and Newton, C. 2015. Twitter CEO: 'We suck at dealing with abuse'. *theverge.com* (Feb. 2015). DOI= http://www.theverge.com/2015/2/4/7982099/twitter-ceo-sent-memo-taking-personal-responsibility-for-the

[34] https://support.twitter.com/articles/20170133-offensive-content

[35] The Institute for Employment Studies. Workplaces and social networking. *ACAS*. DOI= http://www.acas.org.uk/media/pdf/d/6/1111_Workplaces_and_Social_Networking.pdf

[36] https://www.gov.uk/jury-service/discussing-the-trial

[37] Britland, M. 2012. Social media for schools: a guide to Facebook, Twitter and Pinterest. *theguardian.com* (July 2012). DOI= http://www.theguardian.com/teacher-network/2012/jul/26/social-media-teacher-guide

[38] Procter, P., Vis, F. and Voss, A. 2013. Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*. 16,3, (Apr 2013) 197-214. DOI: 10.1080/13645579.2013.774172

[39] Ellis-Peterson, H. 2014. Mary Beard reveals she befriended twitter trolls following online abuse. *theguardian.com* (Aug. 2014). DOI= http://www.theguardian.com/books/2014/aug/27/mary-beard-befriends-twitter-trolls-online-abuse

[40] Nolan, G. 2013. Internet trolls thrive on attention – but please don't feed the animals. *The logical libertarian* (April 2013) DOI= http://logicallibertarian.com/tag/internet-troll/

[41] Carsten Stahl, B. 2012. Morality, Ethics and Reflection: A categorisation of normative research in IS research. *Journal of the Association for Information Systems*. *13*,8, (Aug 2012) 636–656. DOI= http://aisel.aisnet.org/jais/vol13/iss8/1/ ; Carsten Stahl, B. Eden, G. Jirotka, M. and Coeckelbergh, M. 2014. From Computer Ethics to Responsible Research and Innovation in ICT: The transition of reference discourses informing ethics-related research in information systems. *Information & Management*. 51, 6 (Sep 2014) 810-818. DOI= doi:10.1016/j.im.2014.01.001

[42] For more information see our project website www.digitalwildfire.org

Understanding Academic Attitudes Towards the Ethical Challenges Posed by Social Media Research

Chris James Carter[1], Ansgar Koene[1], Elvira Perez[1], Ramona Statache[1], Svenja Adolphs[2], Claire O'Malley[3], Tom Rodden[1], Derek McAuley[1]

[1] Horizon Digital Economy Research Institute, University of Nottingham, [2] School of English, University of Nottingham, [3] Faculty of Science, University of Nottingham Malaysia Campus

{christopher.carter, ansgar.koene, evira.perez, ramona.statache, svenja.adolphs, claire.omalley, tom.rodden, derek.mcauley}@nottingham.ac.uk

ABSTRACT

In this paper, we outline an online survey-based study seeking to understand academic attitudes towards social media research ethics (SMRE). As the exploratory phase of a wider research project, findings are discussed in relation to the responses of 30 participants, spanning multiple faculties and locations at one international university. The paper presents an empirical measure of attitudes towards social media research ethics, reflecting core issues outlined throughout the nascent Internet-mediated research (IMR) literature, in addition to survey questions relating to familiarity with SMRE guidance, and experience of reviewing SMRE proposals from students and/or as part of the university's research ethics committees (RECs). Findings indicate notable variance in academic attitudes towards the ethical challenges of social media research, reflecting the complexity of decision-making within this context and further emphasising the need to understand influencing factors. Future directions are discussed in relation to the tentative findings presented by the current study.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues

General Terms

Measurement

Keywords

Research Ethics, Internet-Mediated Research, Social Media, Research Ethics Committees, Institutional Review Boards, Policy

1. INTRODUCTION

With social media sites such as Facebook and Twitter continuing to attract hundreds of millions of monthly active users [1, 2], the equally vast amount of personal data produced through these services provide academic researchers with unprecedented opportunity for investigating human behaviour online [3]. Analysis of "big data" sets has enabled researchers to explore social phenomena ranging from voting behaviour in elections [4] and self-censorship of status updates prior to posting [5], to the social transference of emotional states [6] and accurate prediction of highly sensitive personal characteristics, such as political and religious affiliations, ethnicity, gender, sexuality, and personality [7, 8].

A steadily expanding body of multidisciplinary research has also adapted various "traditional" research methods such as semi-structured interviews, surveys and participant observation to indirectly explore topics such as motivations in driving social media use [9, 10], including the role of personality [11-14], and the expression of risky behaviour online [15-18]. Through a combination of these two broad methodological approaches, a marked increase has been observed in the number of social media research studies published within the social sciences in recent years, rising from a solitary paper produced in 2005 to a cumulative total of 412 by 2011 upon Facebook [19], and from 3 research papers in 2007 to 527 as of 2011 for Twitter [20].

As the study of social phenomena upon social media continues to increase, so too has the need to understand how academic researchers are addressing the various ethical challenges that are posed by research within this relatively novel environment. Numerous sets of ethical guidelines and recommendations for Internet-mediated research have emerged in recent years [e.g. 21, 22, 23], identifying some of the key ethical issues facing researchers wishing to use social media. However, comparatively little is known about researcher attitudes towards these issues, and how they may translate into experiences of reviewing research ethics proposals submitted by students and fellow academics.

Given the 'bottom-up', researcher-led perspectives adopted within the guidelines published by the Association of Internet Researchers [AoIR: 21, 23] and British Psychological Society [BPS: 22], social media researchers and members of university ethics committees are faced with making challenging, context-specific decisions with respect to judging the ethical appropriateness of Internet-mediated research proposals [3]. Given that members of these ethics review boards may struggle with some of the ethical nuances associated with the emerging field of social media research [24], and in particular studies involving the use of "big data" [25], there is a pressing need to try and understand the attitudes and levels of awareness of academics tasked with this responsibility.

The current paper presents preliminary insights into the attitudes and experiences of a small cohort of academics within a single university, representing the initial piloting phase of a wider study. The following section now turns to provide greater detail on the specific ethical issues presented by social media research, as outline throughout the existing bodies of literature.

Conference'10, Month 1–2, 2010, City, State, Country.

2. RELATED WORK

Initially developed within the context of biomedical research, the core principles of research ethics and the ethical treatment of persons are represented throughout a number of landmark policies and guidelines, including the Nuremberg Code, the Declaration of Helsinki, the National Research Act of 1974, and the Belmont Report. As outlined by Markham and Buchanan [23], "*the basic tenets shared by these policies include the fundamental rights of human dignity, autonomy, protection, safety, maximization of benefits and minimization of harms, or, in the most recent accepted phrasing, respect for persons, justice and beneficence*." (p. 4). These principles are further instantiated through discipline-based guidelines including the Association for Computing Machinery's (ACM) "Code of Ethics and Professional Conduct" [26] and the British Psychological Society's (BPS) "Code of Human Research Ethics" [27], in particular emphasizing the personal and professional responsibilities of researchers.

Following from these sets of codes and principles, universities have implemented Institutional Review Boards (IRBs), or Research Ethics Committees (RECs) in the UK, to review the ethical appropriateness of research study proposals involving human participants within the institution. Indeed, according to the BPS [27], RECs are responsible for ensuring that ethics reviews are conducted in an independent, competent, transparent and timely manner, providing useful feedback and expertise, and ensuring the protection of both researchers and research participants. Despite significant growth in the ethical regulation of research conducted within UK HEIs, and in particular in the social sciences [28, 29], some have questioned the ethics of ethics committees themselves in undermining the freedom and responsibilities of researchers [28], whilst others have argued that humanities and social sciences research simply does not pose the same level of harmful risk as biomedical research [29], thus rendering the extent of ethical regulation in this domain unjustified.

Regardless of the issues inherent to the institutional regulation of research ethics via RECs and IRBs, the increasing prevalence of Internet-mediated research in the last decade is forcing committees to adapt to the unique challenges presented by research within the digital domain. Indeed, ethical decision making is already identified as a complex task [23], but Internet-mediated research introduces further issues and "grey areas" [30] that researchers and ethics review committees may be struggling to adequately engage with. In their review of 30 social media research papers involving young people, Henderson, Johnson, and Auld [24] illustrated this point by finding that only eight articles discussed the ethical challenges associated with their research, and with six of these "*couched in terms of what was required by the university ethics committee, not in terms of ethical considerations or issues arising through the research*" (p. 548). Though the authors stop short of labeling the research as "unethical", they argue that the finding may reflect a limited understanding of social media research-related issues within RECs [24]; a point also echoed by Beaulieu and Estalella [31].

A recent, high profile illustration of this potential issue is provided by the publication of a research study in the Proceedings of the National Academy of Sciences (PNAS) by Kramer, Guillory, and Hancock [6]. Specifically, the research – a joint collaboration between researchers from Facebook, Cornell University and the University of California-San Francisco – used an experimental design to investigate the transference of emotional states on Facebook, covertly manipulating the presentation of status updates conveying positive and negative affect that almost 690,000 users would receive within their profile newsfeed over the period of one week. With the affective basis of the experimental intervention and apparent lack of informed consent, possibility for withdrawal, or debrief, substantial criticism was subsequently aimed at how the study had been granted ethical approval through Cornell University's IRB, with some critics pointing out apparent changes in Facebook's user terms following the study [32] in addition to the aforementioned IRB claimed that they had never reviewed the study, leaving it to Facebook [33].

Though the aforementioned study [6] attracted substantial attention throughout the mainstream press, it is by no means an isolated case of researchers and their respective RECs appearing to underestimate the ethical complexities of social media research. Zimmer [34], for instance, presents a detailed analysis of the numerous ethical issues posed by a study of Facebook user data published by Lewis, Kaufman, Gonzalez, Wimmer, and Christakis [35], titled "Tastes, ties, and time" (T3). In the study, Lewis and colleagues publicly released data collected from the Facebook profiles of 1,700 students sampled across a four-year period at a university in the northeastern United States. Despite attempting to protect the identities of participants by removing names and student identification numbers, and the study receiving ethical approval from the Harvard University IRB, Zimmer [34] was able to successfully breach the anonymity of participants and their institution through combining supplementary aspects of information released in the dataset. Thus, even though the researchers took steps to eliminate privacy violations of the participants' personal data, and that these were deemed sufficient by the university's IRB, ethical issues still remained.

Seeking to outline core issues associated with Internet-mediated research (IMR), the AoIR published their first "Ethical Decision Making and Internet Research" document in 2002 [21]. Rather than drawing upon a top-down approach influenced by the type of principles, regulations, and universal norms outlined previously, Ess and AoIR colleagues' proposal emphasized ethical pluralism, cross-cultural awareness, and a focus on guidelines rather than "recipes"; adopting a more bottom-up stance based upon day-to-day experiences garnered through theoretical, empirical, and field research. Following its application by RECs and IRBs in forming decisions about Internet-mediated research, the AoIR guidelines were subsequently updated by Markham and Buchanan in 2012 to account for more recent developments in the field of IMR, including the subsequent rise of social media [23].

A core point emphasised in this revised proposal [23] continued to be that "*no set of guidelines or rules is static; the fields of Internet research are dynamic and heterogeneous*." (p. 2), and as such, a bottom-up approach to ethical decision-making helped to account for this. In particular, the AoIR guidelines present researchers with a set of considerations to inform the ethical decision-making process, rather than imposing rigid guidance, or hard and fast answers to ethical challenges [23]. This is an important point, as it has clear implications for the requisite knowledge expected of researchers and ethics committee members likely to encounter social media-related research submissions. Specifically, Markham and Buchanan's [23] perspective implies that not only do social media researchers need to possess sufficient awareness of the key principles

guiding ethical research in this domain, but so too do members of the RECs and IRBs tasked with reviewing research proposals of this nature.

A key element of this refers to what the AoIR describe as "major tensions" (p. 6) in IMR, and by extension, social media research. First, the authors identify an ongoing debate about human subjectivity in social media research, or more specifically, whether protocols involving only the indirect involvement of individual users require the same level of ethics committee scrutiny as those that do so more directly. As argued by Beaulieu and Estalella [31], indirect ethnographic research conducted within mediated settings raises distinct ethical issues due to the contiguity and traceability of digital information relating to both researchers and participants. In particular, the authors point out that such issues encourage researchers to consider their accountability towards participants, and that the public nature of online interactions have consequences for the former, as well as the latter.

Relatedly, a second tension posed by the AoIR [23] relates to the status of personhood upon social, and queries whether one's personal data should be considered as an extension of the self, or if it should be treated as a document or text independent of the individual. Indeed, while the value of "small data" detailing rich, lived experiences of individuals upon social media has been emphasized by some over the automated collection of "big data" [25], others have argued that publicly accessible social media content should be treated as documented text, and therefore does not require informed consent from its authors [36].

Additionally, if an aggregated amount of data collected is large enough, the AoIR guidance highlights questions as to the ethical appropriateness of assuming the risk of personal identification is sufficiently reduced. This problematic nature of this assumption has already been introduced with respect to Zimmer's [34] successful de-anonymisation of the T3 research data set [35], in addition to the controversial practice of using verbatim quotes from participants that can potentially be found within public archives of social media data [30]. Indeed, these issues were touched upon in a set of guidelines published in 2007 by the BPS [37]. Specifically, the BPS identified two key dimensions of importance: *level of identifiability* (i.e. from being anonymous, to being identifiable) and *level of observation* (i.e. being covertly observed, through to explicit consent), with various ethical issues subsumed within the subsequent categories, as illustrated in Table 1.

Table 1. BPS [37] typology of four types of IMR studies and examples of ten ethical issues raised

Participants	Identifiable	Anonymous
Recruited	Verifying identity Informed consent Withdrawal Data protection	Levels of control Monitoring the consequences of research Protecting participants and researchers
Unaware	Deception	Understanding of public and private space Debriefing

Reflecting the lower-right quadrant of Table 1, a third tension identified by the AoIR is the public-private distinction, relating to expectations of privacy and whether data shared publicly on social media can indeed be considered as 'private'. As illustrated by the findings of both Henderson *et al* [24] and Weller and Kinder-Kurlanda [30], a number of social media researchers appear to argue against the need for an ethics review to be conducted when data is shared within the public domain, working on the assumption that users are aware of participating in public communication. This issue is also highlighted as a key "ethical dilemma" by Henderson and colleagues [24], who emphasise that participant understanding of private and public online behaviour may be particularly compromised amongst young adults, making the issue even more important for researchers interacting with members of this cohort.

Shifting towards more disciple-based guidelines and building upon the aforementioned set produced in 2007 [37], the BPS' "Ethical Guidelines for Internet-Mediated Research" document [22] further reflects some of the key concerns identified by the AoIR [23]. In particular, the BPS similarly highlight the importance of subjective judgment on the part of the researcher, declaring that the document "*is not intended to provide a 'rule book' for IMR*", and advocating "*a return to 'first principles' and an informed application of general ethics principles to the new situation [of Internet-mediated research]*" (BPS, 2013: 2). In particular, the BPS identifies four core ethical principles for members to adhere to: respect for the autonomy and dignity of persons, including issues relating to the public-private distinction, confidentiality, copyright, valid consent, withdrawal, and debriefing; scientific value; social responsibility; and maximizing benefits and minimizing harm.

In recent years, a number of UK-based research groups have emerged within universities to examine the ethical issues associated with social media analysis. For instance, the ESRC-funded Collaborative Online Social Media Observatory (COSMOS) [38] and Citizen-centric Approaches to Social Media Analysis (CaSMa)[1] [39] research groups based at Cardiff University and the University of Nottingham, respectively, both adopt clear, person-centered and ethically rigorous approaches to the design of social media research studies. However, with researchers and RECs within universities faced with forming decisions that balance the rights of human participants against the social benefits of research proposals, it is not yet clear as to how aligned these groups are with the perspectives of COSMOS, CaSMa [39], and similar research groups.

As discussed in outlining the predominantly "bottom-up" perspectives of some of the most comprehensive existing IMR guidelines [22, 23], a key characteristic appears to be in advocating pragmatic and responsible decision making on the part of the researcher. As remarked by Markham and Buchanan [23], this appears to reflect that "*there is much grey area in ethical decision-making … Multiple judgments are possible, and ambiguity and uncertainty are part of the process*" (p. 5). With social media adding to the complexity of ethical decision making [30], and research ethics committees seemingly struggling with this [24, 25], the main research questions addressed by the study presented in the current paper were,

RQ1: *How do academics tasked with ethically reviewing research proposals perceive the ethical challenges posed by social media research?*

And additionally,

[1] CaSMa is the Horizon Digital Economy Research Institute group that the authors of the current paper are affiliated with.

RQ2: *How do attitudes towards social media research ethics (SMRE) relate to experience of reviewing research proposals of this type, and experience of Internet-mediated ethical guidelines and training?*

The following section now outlines the findings of a small empirical study of academic attitudes towards SMRE, conducted as the piloting phase of a larger research project to unfold across the coming months.

3. DESIGN

3.1 Participants

Participants were 30 academic members of staff employed by a Russell Group university, with the majority of respondents based on the institute's UK-based campuses (n = 20, 74.1%) and the remainder located internationally (n = 7, 25.9%; n = 3 undeclared). Participants responded to an email request containing a hyperlink to an online survey, sent via the respective Heads of the institution's 26 School Ethics Committees. The sample comprised of 18 males (64.3%) and 10 females (35.7%; n = 2 undeclared). The median and modal age band of participants was 35 to 44-years-old. All five faculties at the institution were represented in the sample, though particularly Science (n = 9, 32.1%), Social Sciences (n = 5, 17.9%), and Medicine and Health Sciences (n = 5, 17.9%).

3.2 Measures

The online survey used in the study consisted of basic demographic questions (e.g. age, gender, location, current faculty) in addition to three sections of questions measuring experience of reviewing social media research ethics (SMRE) proposals at the institution, experience of SMRE guidance and training, and attitudes towards SMRE. These sections are now described in more detail in the following sub-sections.

3.2.1 Experience of Reviewing Social Media Research Ethics Proposals

For participants indicating that they held the responsibility of reviewing student research ethics proposals, and/or were members of their School Ethics Committee, the online survey asked whether they had experience of reviewing research ethics proposals involving the use of social media, indicating either *Yes, No, or Other*. Participants were also asked how they would describe their level of confidence in being able to identify ethical issues specifically related to social media research proposals, using a 5-point Likert scale anchored at 1 (*Not at all confident*) and 5 (*Extremely confident*). Participants were also asked how they would describe their experience of reviewing SMRE proposals in relation to "traditional" proposals relating to offline behaviour, using a 5-point Likert scale anchored at 1 (*Significantly easier than reviewing "traditional submissions*) and 5 (*Significantly harder than reviewing "traditional" submissions*).

3.2.2 Experience of Social Media Research Ethics Guidance and Training

Participants were asked whether they had received any formal training or guidance from their institution in dealing with ethically reviewing social media research proposals, indicating either *Yes, No, or Other*. The survey also asked participants to indicate whether they were familiar (*Yes/No/Other*) with a number of research ethics documents including their institution's code of research conduct and research ethics document, its specific "e-Ethics" guidance document, the AoIR's [23] "Ethical Decision-Making and Internet Research" document, and any Internet-mediated research guidelines produced by their specific academic discipline, such as the BPS [22] and ACM [26]. If answering "*Yes*", participants were asked how useful they found the documents in providing guidance for reviewing social media research proposals, using a 5-point Likert scale anchored at 1 (*Not at all useful*) and 5 (*Extremely useful*).

3.2.3 Attitudes Towards Social Media Research Ethics

In order to measure attitudes towards SMRE, a pool of 12 items was developed that would reflect some of the core ethical issues discussed previously in Section 2. Specifically, 12 statements were constructed, and to be measured using a 7-point Likert scale anchored at 1 (*Strongly disagree*) and 7 (*Strongly agree*), and with a neutral mid-point at 4 (*Neither agree nor disagree*). The specific wording of these statements is found in Table 2, with participants asked to indicate their level of agreement with each using the scale provided. Ethical issues covered by the statements included attitudes towards gaining informed consent (Q1, Q2, Q4, Q7, Q11), the public-private distinction (Q1, Q6, Q7, Q8), anonymity (Q3), withdrawal (Q2), personhood (Q10), and deception (Q12), in addition to more general attitudes towards the relative costs and benefits of ethical decision making when doing social media research (Q4, Q5, Q9).

With the exception of Q2 ("*Individuals must always be informed of their participation in social media research so that they may withdraw from the study*"), all remaining statements were designed so that disagreement (i.e. low scores) would reflect the type of person-centred, ethically-driven attititudes towards social media research adopted by researchers [3, 25, 30] and research groups, such as CaSMa and COSMOS. Though Table 2 presents these statements in their original direction, the composite measure of attitudes towards social media ethics presented in the Results section reversed all items other than Q2, so that higher overall scores would represent greater alignment with the aforementioned person-centred, ethically-driven attitudes towards social media research.

Table 2. Attitudes Towards Social Media Research Ethics – Item Descriptions

Items	Item Description
Q1	"There is no need to gain informed consent to do research with an individual's social media data if it is publicly accessible"
Q2	"Individuals must always be informed of their participation in social media research so that they may withdraw from the study"
Q3	"It is very unlikely that individuals will be able to be identified if social media datasets are anonymised"
Q4	"Seeking informed consent from individuals unknowingly involved in social media research typically creates more problems for researchers than are necessary"
Q5	"It is too impractical to expect researchers to apply every ethical consideration associated with human research to studies using social media data"
Q6	"It is the responsibility of individuals to rethink how they use social media if they are unwilling

	for their online public behaviour to be studied by researchers"
Q7	"It is acceptable for researchers to use publicly accessible data on social media without prior informed consent of the individuals who published it"
Q8	"There is no discernible ethical difference between studying the public behaviour of individuals on social media to those in real world public settings"
Q9	"The beneficial outcomes of being able to study human behaviour through social media data typically outweigh the need to inform users of their participation"
Q10	"Studying the publicly accessible social media data of individuals is essentially equivalent to researching document-based text, where human research ethics do not apply"
Q11	"Agreement with the 'terms and conditions' of social media sites is sufficient permission for researchers to use data without seeking further consent from users"
Q12	"It would typically be acceptable to provide misleading information about the true purpose of a research study using social media data, so long as the individual was informed at a later stage"

3.3 Procedure

Following approval from the relevant Research Ethics Committee associated with the current authors, the lead author sent an invitation email containing details of the study and a hyperlink to the information page of the online survey to Heads of the 26 faculty-based School Ethics Committees throughout the university involved in the research. Specifically, Heads were asked to disseminate the details of the study to academic colleagues upon their School's ethics committee and/or with the responsibility of reviewing the ethics of undergraduate and/or postgraduate research proposals. Hosted upon the Bristol Online Surveys (BOS) platform, the survey was anonymous, password-protected, and accessed only by the lead author. Both anonymity and withdrawal from the study were ensured by asking participants to provide a unique identifier that could later be quoted, combining their mother's maiden name with the current time of survey completion (e.g. LISTER1045).

Following the provision of consent, participants were first presented with a brief overview of the various types of social media, based upon the typology proposed by Kaplan and Haenlein [40]. They were then shown a brief section outlining different types of social media research based upon the "What is Internet Research?" section on page 3 of the AoIR's 2012 guidelines [23]. Participants were presented first with the 12 items measuring attitudes towards SMRE (see Section 3.2.3), followed by questions relating to experience of reviewing SMRE proposals (see Section 3.2.1), and then experience of SMRE guidance and training (see Section 3.2.2). The survey closed with a section asking basic demographic questions (see Section 3.2) and providing debriefing materials about the study, including a link to further information about the research, hosted upon the CaSMa research blog [39].

4. RESULTS

The majority of participants reported holding the responsibility of reviewing undergraduate and/or postgraduate research ethics proposals (n = 26, 86.6%). Respondents indicated a wide range of experience, from less than 1 year to more than 10 years, resulting in a median and modal experience of 2 to 3 years in the role (29.2%). Within this role, over two-thirds (70.8%, n = 17) reported having ethically reviewed student research proposals that involved the use of social media. Of this sub-group, almost one-third (31.3%, n = 5) reported feeling "very confident" about identifying SMRE issues, with a median and modal response of feeling "moderately confident" (50%, n = 8). No participants indicated being "not at all confident". While just over one-third (37.5%, n = 6) reported that "there was no noticeable difference between reviewing 'traditional' and social media-related submissions", the modal and median response indicated that precisely half found SMRE proposals "slightly harder" (50%, n = 8).

Just over half of the participants reported reviewing research ethics proposals as a member of their School's Research Ethics Committee (56.7%, n = 17), with experience ranging from less than one year to 4 to 5 years, and a median and modal experience of 2 to 3 years in the role (35.3%). Just over three-quarters (76.5%, n = 13) of respondents in this role reported having ethically reviewed research proposals involving the use of social media. Of this subset, one-third (33.3%, n = 4) again reported feeling "very confident" about identifying SMRE issues, whilst the median and modal response was feeling "moderately confident" (58.3%, n = 7). As before, no respondents indicated feeling no confidence at all. Though one-third (33.3%, n = 4) reported that "there was no noticeable difference between reviewing 'traditional' and social media-related submissions", the modal and median response indicated that almost three-fifths found SMRE proposals "slightly harder" (58.3%, n = 7).

Precisely four-fifths (80%, n = 24) of respondents indicated having never received formal training or guidance on handling SMRE proposals, with the remaining one-fifth (20%, n = 6) having done so through general ethics training from their university, workshop-based discussions, and through attending presentations and reading articles. Almost all participants reported being familiar with the university's code of research conduct and research ethics document (96.7%, n = 29), with the majority of respondents finding it "moderately useful" (44.8%, n = 13) in providing guidance for reviewing SMRE proposals (mean = 2.76; S.D. = 1.02; median and mode = 3).

Familiarity with the university's specific e-ethics document was more balanced, with only just over half (52%, n = 13) indicating an awareness of it. Of this subset, just over half (53.8%, n = 7) found it "moderately useful" in providing guidance for reviewing SMRE proposals (mean = 3.31, S.D. = .63), though almost two-fifths also reported it as "very useful" (38.5%, n = 5). Relatively few respondents were familiar with either the AoIR [23] guidance (16.7%, n = 5) or their own academic discipline's IMR guidelines (26.7%, n = 8).

A number of interesting findings are indicated in Table 3, where the means and standard deviations of responses to each of the 12 Attitudes Towards Social Media Research Ethics (SMRE) items are presented, along with composite levels of disagreement and agreement (slightly, moderately, and strongly combined).

Table 3. Attitudes Towards Social Media Research Ethics - Means, Standard Deviations, and Agreement (in %)

Items	Item Description		
Mean (SD)	*Disagree*	*Neither Ag. nor Dis.*	*Agree*
Q1	No need for informed consent if SM data publicly accessible		
3.53 (2.19)	*60%*	*3.4%*	*36.6%*
Q2	Informed consent required to enable withdrawal from SM research		
4.47 (2.27)	*40%*	*3.4%*	*56.6%*
Q3	Unlikely that individuals will be identified if SM dataset is anonymous		
3.67 (1.81)	*63.3%*	*6.7%*	*30%*
Q4	Informed consent creates more problems for SM researchers than necessary		
4.17 (1.66)	*27.6%*	*31%*	*41.4%*
Q5	Too impractical to apply all ethical considerations to SM research		
3.47 (1.80)	*50%*	*13.3%*	*36.7%*
Q6	Responsibility is upon individuals if they do not wish to participate in SM research		
4.37 (2.21)	*43.4%*	*0*	*56.6%*
Q7	Acceptable to use public SM data without informed consent		
4.33 (2.01)	*43.4%*	*0*	*56.6%*
Q8	No ethical difference between studying offline and SM behaviour in public spaces		
4.10 (1.97)	*46.6%*	*10%*	*43.4%*
Q9	Benefits of studying behaviour on SM outweigh need for informed consent		
2.97 (1.59)	*60%*	*26.6%*	*13.4%*
Q10	Studying public data on SM is essentially same as studying documented text		
2.97 (1.96)	*73.3%*	*3.4%*	*23.3%*
Q11	User agreement with SM terms and conditions sufficient as informed consent		
3.13 (2.01)	*60%*	*13.4%*	*26.6%*
Q12	Acceptable to deceive SM users in research as long as informed at a later date		
2.63 (1.56)	*73.3%*	*13.3%*	*13.4%*

Many of the responses to items present a complex picture in which respondents appeared to recognise the ethical importance of avoiding deception (Q12) and gaining consent from participants in social media research (Q1, Q2, Q9, and Q11), but also seemed to acknowledge the increased problems facing researchers in doing so (Q4).

Similarly, most respondents disagreed to some extent with the notion that studying public data upon social media was essentially the same as studying documented text (Q10: 73.%) and that individuals wouldn't be identified from large datasets if anonymous (Q3: 63.3%), yet levels of agreement and disagreement were roughly equivocal with respect to the acceptability of using such data without informed consent (Q7), the ethical equivalence of researching in offline and online public spaces (Q8), and the responsibility of users in indicating willingness to participate (Q6).

With standard deviations for each of the 12 Attitudes Towards SMRE items ranging from 1.56 (Q12) to 2.27 (Q2), there appeared to be considerable variance across the responses. Though the restricted sample size meant that exploratory factor analysis was inappropriate as a means of investigating the relationships between items, inter-item correlations were calculated to examine whether statistically significant positive relationships could be found to indicate the measurement of one or more constructs. For 10 of the 12 items, item-total correlations ranged from r = .465 (Q5) to r = .804 (Q10), though the two items of Q8 and Q12 appeared to exhibit notably different item-total correlations of r = -.121 and r = .080, respectively. Further inspection of the correlation matrix confirmed that Q8 featured only one statistically significant relationship with the remaining 11 items (Q5: r = -.384, $p < .05$), and Q12 shared none.

Reliability analysis revealed that Cronbach's Alpha improved from α = .837 for all 12 items, to a good internal consistency of α = .889 when removing Q8 and Q12 to form a 10-item composite measure. The mean score for the resulting measure was 4.39, with a standard deviation of 1.38. To explore the second research question underpinning the study (see RQ2, Section 2), one-way analyses of variance (ANOVAs) were conducted and found no significant differences in scores on the Attitudes Towards SMRE items based on experience of having reviewed SMRE proposals submitted by students ($F(1,22) = 3.51$, $p = .074$, n.s.) or as part of their role upon the school ethics committee ($F(1,15) = .27$, $p = .612$, n.s.).

Similarly, no significant differences were found based on experience of formal SMRE training or guidance ($F(1,28) = 2.12$, $p = .157$, n.s.) or familiarity with the university's e-ethics document ($F(1,23) = 2.05$, $p = .166$, n.s.), the AoIR's IMR guidelines ($F(1,28) = 0.05$, $p = .827$, n.s.), or any IMR guidance provided by their academic discipline ($F(1,28) = 1.24$, $p = .275$, n.s.). Correlational analyses also revealed statistically non-significant relationships between Attitudes Towards SMRE scores and level of experience in reviewing student research ethics proposals (r = .09, $p = .69$, n = 24, n.s.) and reviewing as part of the school ethics committee (r = .09, $p = .73$, n = 17, n.s.). The relationship with level of confidence in being able to identity SMRE issues in both student (r = .19, $p = .49$, n = 16, n.s.) and REC submissions (r = .04, $p = .89$, n = 12, n.s.) was also found to lack statistical significance, although this is not unexpected given the particularly restricted sample sizes involved.

5. DISCUSSION

The current paper has outlined the findings of an initial, exploratory phase of a wider research project investigating academic attitudes towards social media research ethics (SMRE). Though the limited number (n = 30) of respondents and single institutional source from which participants were sampled significantly restrict the generalisability of the findings, the study nevertheless provides the foundations for a crucial - albeit tentative - discussion of the empirical study of social media research ethics. Indeed, reflecting the apparent rise in

academic research involving social media [19, 20], the study found evidence indicating that most respondents had reviewed an SMRE proposal, whether submitted by undergraduates and postgraduates under their supervision, or as a member of their school's research ethics committee (REC).

With respect to the first research question of how academics tasked with ethically reviewing research proposals perceive the ethical challenges posed by social media research, the study produced a number of interesting findings. For instance, despite the apparent prevalence of social media research submitted for review within the university, relatively few respondents reported having received any formal training or guidance in reviewing research proposals of this nature. Nevertheless, just over two-fifths found their university's general research ethics guidance to be moderately useful in doing so, while just over half were familiar with their institution's "e-ethics" research guidelines, which were also found to be largely helpful. In contrast, relatively few respondents reported being familiar with the comprehensive AoIR guidelines [23] or discipline-based Internet-mediated Research (IMR) guidance exemplified by the BPS [22], and outlined in previously in Section 2.

In terms of attitudes towards some of the core ethical challenges of social media research, as outlined in the aforementioned guidelines and discussed by the likes of Henderson and colleagues [24] and Moreno *et al* [3], a number of interesting points are apparent. In particular, a majority of respondents appeared to indicate an understanding of the need for informed consent and avoidance of deception when doing social media research, in addition to an appreciation that online data may not simply be regarded as text-based documents [cf. 41] and that large, anonymous datasets do not rule out potential violations of participant privacy, as demonstrated by Zimmer [34] in relation to the "T3" study [35]. In respect to these issues, many respondents seemed to convey attitudes aligned with the person-centred perspectives adopted by the likes of the COSMOS [38] and CaSMa [39] research groups described in Section 2.

However, attitudes appeared more balanced across the sample with respect to other issues. In particular, similar proportions of agreement and disagreement were found in relation to whether public data necessitates the need for informed consent, whether there are any fundamental differences between studying offline and online public behaviour, and whether seeking informed consent may create more problems for researchers than necessary. The relatively large standard deviations of responses suggest notable variation in attitudes across the sample, and indeed, this may be expected given the complexity of the issues [3, 23] and the broad range of disciplines included in the otherwise limited sample frame. This level of complexity is also reflected in evidence suggesting that many academics find reviewing SMRE proposals slightly more difficult than 'traditional' research proposals within an offline context, though nevertheless remain moderately confident about their ability to successfully detect ethical issues specific to IMR.

With regards to the second research question, no statistically significant relationships were found between attitudes towards SMRE and experience of reviewing research proposals of this type, or experience of IMR ethical guidelines and training. Though no specific hypotheses were offered in the current study, it might have been expected that experience of reviewing social media proposals, attendance of formal SMRE training, or familiarity with SMRE guidelines and principals would be positively related to more person-centred attitudes. In fact, the test closest to reaching statistical significance indicated greater scores on the attitudes to SMRE scale being reported by respondents with *no* experience of reviewing student social media research proposals compared to those who had (mean = 5.04 vs. 3.93), hinting towards the possibility that the idealistic principles of the person-centred approach to social media research ethics may reduce when presented with the many complexities of practical experience. Given the restricted sample size, however, this possibility would need to be examined further in future studies.

For similar reasons, the study was unable to explore the psychometric structure of the 12 items measuring attitudes towards SMRE, and therefore, whether they represent a single construct (e.g. a person-centred approach to social media research ethics) or multiple facets. However, despite this limitation, reliability analysis and close inspection of the inter-item correlation matrix enabled the identification of two problematic items which, unlike the remaining 10 items which all positively correlated with one another, failed to significantly do so in more than one instance. Following their removal, the subsequent 10-item scale demonstrated very good internal consistency ($\alpha = .89$), which provides a promising foundation for further testing and use of the items as an empirical measure of attitudes towards SMRE in future research. Indeed, it is in this direction that future research conducted by the CaSMa research group is to turn, following on from the initial exploratory phase presented in this paper.

In particular, one forthcoming study will use semi-structured interviews to gain greater depth of understanding in attitudes towards SMRE and the apparent gap between familiarity with IMR ethical guidelines and confidence in addressing related issues, building upon both the present study and recent work by Weller and Kinder-Kurlanda [30]. A further study using a revised version of the current online survey will be made accessible to stakeholders across multiple institutions, thus widening the breadth of the sample and enabling greater statistical power to explore some of the relationships proposed, and tentatively addressed in the current study.

Despite a range of comprehensive guidelines and authors interested in social computing increasingly turning their attention towards the ethical challenges posed by the increasingly popular field of social media research, the ways in which academics tasked with integrating these considerations into ethical decision-making do so on a practical basis is still, as yet, relatively unclear. Complementing theoretical work in this area with empirical research seems likely to provide exciting opportunities for better understanding the nuances of ethical decision-making in designing and evaluating social media researching. It is hoped that the current paper will provide a suitable platform from which such discussions and research can continue to flourish.

6. ACKNOWLEDGMENTS

This work forms part of the Citizen-centric Approaches to Social Media Analysis (CaSMa) project, supported by ESRC grant ES/M00161X/1 and based at the Horizon Digital Economy Research Institute, University of Nottingham. For more information about the CaSMa project, please see http://casma.wp.horizon.ac.uk/.

7. REFERENCES

[1] Facebook. 2015. *Statistics*. Available at www.newsroom.fb.com/company-info/

[2] Twitter. 2015. *Twitter Usage*. Available at www.about.twitter.com/company

[3] Moreno, M. A., Goniu, N., Moreno, P. S. and Diekema, D. 2013. Ethics of social media research: common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking*, *16*, 9, 708-713.

[4] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E. and Fowler, J. H. 2012. A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 7415, 295-298.

[5] Das, S. and Kramer, A. 2013. Self-Censorship on Facebook, in *Proceedings on the 7th Annual AAAI Conference on Weblogs and Social Media* (Cambridge, Massachussetts, July 2013) AAAI Press, 120-127.

[6] Kramer, A. D., Guillory, J. E. and Hancock, J. T. 2014. Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks, *Proceedings of the National Academy of Sciences*, *111*, 24, 8788-8790.

[7] Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P. and Stillwell, D. 2012. Personality and Patterns of Facebook Usage, in *Proceedings of the ACM Web Science Conference* (Evanston, Illinois, June 2012) ACM New York, 36-44.

[8] Kosinski, M., Stillwell, D. and Graepel, T. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*, 15, 5802-5805.

[9] Nadkarni, A. and Hofmann, S. G. 2012. Why Do People Use Facebook? *Personality and Individual Differences*, *52*, 3, 243-249.

[10] Seidman, G. 2012. Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, *54,* 3, 402-407.

[11] Amichai-Hamburger, Y. and Vinitzky, G. 2010. Social network use and personality. *Computers in Human Behavior*, *26*, 6, 1289-1295.

[12] Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B. and Gosling, S. D. 2010. Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*, 3, 372-374.

[13] Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N. and Gaddis, S. 2011. Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, *14*, 9, 483-488.

[14] Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G. and Orr, R. R. 2009. Personality and motivations associated with Facebook use. *Computers in Human Behavior*, *25*, 2, 578-586.

[15] Sleeper, M., Cranshaw, J., Kelley, P. G., Ur, B., Acquisti, A., Cranor, L. F. and Sadeh, N. 2013. I read my Twitter the next morning and was astonished: A conversational perspective on Twitter regrets, in *Proceedings of the 2013 ACM Annual Conference on Human Factors in Computing Systems* (Paris, France, April 2013) ACM New York, 3277-3286.

[16] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G. and Cranor, L. F. 2011. I regretted the minute I pressed share: A qualitative study of regrets on Facebook, in *Proceedings of the 7th Symposium on Usable Privacy and Security* (Pittsburg, Philadelphia, July 2011) ACM New York, 1-13.

[17] Karl, K., Peluchette, J. and Schlaegel, C. 2010. Who's posting Facebook faux pas? A cross‐cultural examination of personality differences. *International Journal of Selection and Assessment*, *18*, 2, 174-186.

[18] Peluchette, J. and Karl, K. 2009. Examining students' intended image on Facebook: "What were they thinking?!". *Journal of Education for Business*, *85*, 1, 30-37.

[19] Wilson, R. E., Gosling, S. D. and Graham, L. T. 2012. A Review of Facebook Research in the Social Sciences. *Perspectives on Psychological Science*, *7*, 3, 203-220.

[20] Williams, S. A., Terras, M. M. and Warwick, C. 2013. What do people study when they study Twitter? Classifying Twitter related academic papers. *Journal of Documentation*, *69*, 3, 384-410.

[21] Ess, C. 2002. *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee*. Available at http://aoir.org/reports/ethics2.pdf

[22] British Psychological Society. 2013. *Ethics Guidelines for Internet-mediated Research*. British Psychological Society, Leicester: UK. Available at http://www.bps.org.uk/system/files/Public%20files/inf206-guidelines-for-internet-mediated-research.pdf

[23] Markham, A. and Buchanan, E. 2012. *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee (Version 2.0)*. Available at http://aoir.org/reports/ethics2.pdf

[24] Henderson, M., Johnson, N. F. and Auld, G. 2013. Silences of ethical practice: Dilemmas for researchers using social media. *Educational research and evaluation*, *19*, 6, 546-560.

[25] boyd, d. and Crawford, K. 2011. Six Provocations for Big Data, in *Proceedings of the A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society* (Oxford Internet Institute, Oxford, September 2011), 1-17.

[26] Anderson, R. E. 1992. ACM code of ethics and professional conduct. *Communications of the ACM*, *35*, 5, 94-99.

[27] British Psychological Society. 2014. *Code of Human Research Ethics*. Leicester: UK. Available at http://www.bps.org.uk/sites/default/files/documents/code_of_human_research_ethics.pdf

[28] Hammersley, M. 2009. Against the ethicists: on the evils of ethical regulation. *International Journal of Social Research Methodology*, *12*, 3, 211-225.

[29] Dingwall, R. 2008. The ethical case against ethical regulation in humanities and social science research. *Twenty-First Century Society: Journal of the Academy of Social Sciences*, *3*, 1, 1-12.

[30] Weller, K. and Kinder-Kurlanda, K. 2014. "I love thinking about ethics!" Perspectives on ethics in social media research, in *Selected Papers of Internet Research* (Deagu, South Korea, Oct 2014). Available at https://katrinweller.files.wordpress.com/2012/08/hiddendataethics_wellerkinder-kurlanda_ir15-preprint.pdf

[31] Beaulieu, A. and Estalella, A. 2012. Rethinking research ethics for mediated settings. *Information, Communication & Society*, *15*, 1, 23-42.

[32] Hill, K. 2014 (1 July). *Facebook Added 'Research' To User Agreement 4 Months After the Emotion Manipulation Study*, Forbes, Available at http://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/.

[33] Hill, K. 2014 (29 June). *Facebook Doesn't Understand The Fuss About Its Emotion Manipulation Study*, Forbes, Available at http://www.forbes.com/sites/kashmirhill/2014/06/29/facebook-doesnt-understand-the-fuss-about-its-emotion-manipulation-study/.

[34] Zimmer, M. 2010. "But the data is already public": on the ethics of research in Facebook. *Ethics and Information Technology*, *12*, 4, 313-325.

[35] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. and Christakis, N. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, *30*, 4, 330-342.

[36] Wilkinson, D. and Thelwall, M. 2011. Researching Personal Information on the Public Web: Methods and Ethics. *Social Science Computer Review*, *29*, 4, 387-401.

[37] British Psychological Society. 2007. *Guidelines for ethical practice in psychological research online*. British Psychological Society, Leicester: UK. Available at http://www.bps.org.uk/sites/default/files/documents/conducting_research_on_the_internet-guidelines_for_ethical_practice_in_psychological_research_online.pdf.

[38] Collaborative Online Social Media Observatory (COSMOS). 2015. *What is COSMOS?* Available at www.cs.cf.ac.uk/cosmos.

[39] Citizen-centric Approaches to Social Media Analysis (CaSMa). 2015. *CaSMa Research*. Available at www.casma.wp.horizon.ac.uk/.

[40] Kaplan, A. M. and Haenlein, M. 2010. Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*, 59-68.

[41] Wilkinson, D. and Thelwall, M. 2011. Researching personal information on the public web methods and ethics. *Social Science Computer Review*, *29*, 4, 387-401.

The Path Dependence of Dynamic Traditions and the Illusion of Cultural AIDS

Richard Volkman
Southern Connecticut State University
501 Crescent Street
New Haven, CT 06514
1-203-392-6780
volkmanr1@southernct.edu

ABSTRACT
Analyses of cultural change routinely turn on observations or evaluations regarding what some institution, system of belief, or technology is doing to "us," but it can be obscure how one is supposed to fix the meaning of such claims. This essay explores such analyses, calling attention to their reliance on the rhetorical force of the first person plural when the literal meaning of their claims strongly suggests the third person would be more literally appropriate. In many cases, "we" has to mean "them—not us." The essay describes how this rhetorical move invites readers to conceive the relation between individuals and the cultures they inhabit as legitimizing a dubious paternalism, how pluralism undermines confidence in the paternalist attitude entwined within that conception, and finally sketches an alternative in which individuals are vested with ultimate cultural authority.

Categories and Subject Descriptors
K.4.1 [**Public Policy Issues**]: Ethics

General Terms
Human Factors

Keywords
Postman, Emerson, culture, intuition pump, philosophy, pluralism

1. INTRODUCTION
Analyses of cultural change routinely turn on observations or evaluations regarding what some institution, system of belief, or technology is doing to "us," but it can be obscure how one is supposed to fix the meaning of such claims. Who or what is the referent of the collective pronouns in such claims, and how exactly do such claims translate into reasons for action at the level of individual decisions? What does a change in the way *we* think mean for *me* and *you*? It cannot be doubted that such claims are routinely defended and taken seriously. There is no shortage of celebrated articles, stretching back to the dawn of the computer revolution, imploring "us" to examine the impact of technology on "our" culture and what "we" can and should do about it. Nicholas Carr (2008) wonders "Is Google Making Us Stupid?" Sherry Turkle (2004) promises to tell us "How Computers Change the Way We Think," and Neil Postman (1990) warns against "Informing Ourselves to Death."

With an explicit awareness of the self-referential nature of my thesis, I argue that we need to be careful whenever we argue that the computer revolution changes the way we think, since it is seldom obvious who "we" are. I argue that cultural criticism routinely slips from being about "us" to being about "them," and this slippage obscures a dubious framing of the relation between cultures and individuals that begets dire illusions with respect to our cultural circumstance.

The essay proceeds by unpacking and evaluating the claim by Postman that "our defenses against information glut have broken down; our information immune system is inoperable. We don't know how to filter it out; we don't know how to reduce it; we don't know to use it. We suffer from a kind of cultural AIDS" [7]. The claim that we suffer from cultural AIDS cannot be rigorously defended by logic or statistics, and this is not the strategy Postman adopts. Rather, the story of our contemporary cultural circumstance as characterized by cultural AIDS is a kind of "intuition pump," in which the reader is brought to *see* things as the author sees them by being invited to think through the author's thoughts. This leads to a paradox: since those afflicted with cultural AIDS do not have the resources to judge for themselves whether or not they have cultural AIDs, when Postman declares that "we" suffer from cultural AIDS, he has to mean *they* suffer from cultural AIDS. Those who can understand and appreciate his view, including Postman himself, manifestly do have the resources to filter, reduce, and use the information at their disposal.

Since claims about what the information age is doing to us are often disguised worries about what it is doing to *others*, it must be asked whether we are interested and not unbiased observers of their plight. Indeed, one has to wonder whether the purported observation that our culture is afflicted with a cultural AIDS is not an illusion borne of the inability to accurately peer into the sources of meaning and purpose that inform the navigation of those whose lives and interests are alien to us. If individuals are managing to find their ways in the information age by means Postman and his ilk cannot fully understand let alone endorse, then it is no surprise when they look at the culture and see only a barren nihilism issuing in cultural AIDS. There is another way to see this landscape, which reimagines the relation between individuals and the flourishing cultures they create.

2. CULTURAL AIDS ON THE MOON
In *Technopoly*, Postman pursues in greater depth the story he started telling in his 1990 address to the German Informatics Society, "Informing ourselves to death," whose title plays on Postman's 1985 discussion of television culture, *Amusing*

Conference'10, Month 1–2, 2010, City, State, Country.

Ourselves to Death. Although the ideas pursued in these works resist being boiled down overmuch, the main idea running through them is clear enough: Contemporary media technologies spawn a culture in which individuals are encouraged to immerse themselves in various shallow and fleeting experiences to the detriment of their abilities to form deep and meaningful notions of themselves; they are "amused" or "informed" in their encounters with the culture instead of being edified or provoked to make something of themselves by reflecting on the ultimate nature of the World and their place in it. In the "technopoly" of contemporary America, by which Postman means the state of culture and its corresponding state of mind wherein the logic of our technologies enjoys a monopoly over our thinking, efficiency and scientism supplant other values and commitments to more transcendent ideals. In this story, technopoly is a form of "cultural AIDS."

Postman explains, "All societies have institutions and techniques that function as does a biological immune system. Their purpose is to maintain a balance between the old and the new, between novelty and tradition, between meaning and conceptual disorder, and they do so by 'destroying' unwanted information" [8, 72-3]. Starting with the advent of printing and especially accelerating in the technological milieu of the 19th Century, Postman contends we have been bombarded with information from all directions, undermining any sense of conviction regarding who we are and what matters. "The thrust of a century of scholarship had the effect of making us lose confidence in our belief systems and therefore in ourselves" [8, 55]. Having lost any sense of where we stand and what we stand for, information appeals to us indiscriminately and promiscuously, until we are no longer in a position to make any sense of ourselves and the world we occupy. "The world in which we live is very nearly incomprehensible to most of us. There is almost no fact - whether actual or imagined - that will surprise us for very long, since we have no comprehensive and consistent picture of the world which would make the fact appear as an unacceptable contradiction" [7].

The notion is vividly illustrated in an anecdote my undergraduate literature professor, Emilio DeGrazia, liked to tell students of Classical Mythology. He was visiting relatives in isolated and rural Italy at the time of the moon landing in 1969. Watching the spectacle on television like so many across the globe, he could not help but marvel out loud at the magnificence of this great accomplishment. It was, he opined, a supreme triumph of humanity. To think, right now there are *men* on the *moon*. His elderly great-aunt scoffed at this. He was at first shocked and even a little outraged that she would not regard this event as among the greatest moments of human history, but he was dumbstruck to learn she did not mean to dispute the significance of the achievement but the very fact of it. She did not believe there were men on the moon. But we just saw it on TV! It's been in planning and in the news for years! Hundreds of people participated, and hundreds of thousands more observed the event! She was unruffled: "There are not any men on the moon. There *cannot* be any men on the moon. If there were men on the moon, then where would God live?"

DeGrazia told this story to illustrate for us modern students of ancient religious texts the same basic lesson that Postman means to teach us: Being committed to a worldview of transcendent significance and authority grounds identity and confers meaning in part by eliminating from consideration a certain range of thoughts and beliefs, and the ways of life defined by such commitments are somewhat alien to the modern mind. Where this old woman enjoys a firm certainty, we moderns suffer an identity crisis. Of course, the truth of her beliefs and the worldview that informs them must be judged dubious at best. There *were* men on the moon, after all. Where Postman worries we suffer from cultural AIDS, making us gullible to the extent that we have no point of view and no real identity at all, premodern faith may express hyperactivity of the cultural immune system. But such a casual dismissal of her worldview betrays exactly the sort of misplaced priorities alleged of modernity. Her various mistakes regarding trivial matters of worldly fact, like the locations in space and time of particular astronauts, might be at least somewhat compensated for by her possession of deeper and more significant truths. How she differs from us is misunderstood if we take it to be a difference regarding banal facts. To be sure, she did not mean that God literally lived on the moon and would have to vacate the premises before humans could occupy it; rather, her way of seeing the world is constituted by a sharp distinction between sacred and profane, God and human, heavens and earth that does not admit even the bare conceivability that profane humans might leave earth to visit the sacred heavens of God. To persuade her to change her opinion, one must appeal to matters of greater significance than what we all might think we saw on TV. Our appeal will have to be to what is highest and most worthy of respect. We shall have to move to a discourse beyond mere information mongering and scientism.

None of this implies that we moderns can or should wish to swap our ways of thinking for hers. But the story does bring out the sense in which Postman means to represent technology as a Faustian bargain. Although we gain a great deal in the course of technological change, there is inevitably something lost. "We no longer have a coherent conception of ourselves, and our universe, and our relation to one another and our world. We no longer know, as the Middle Ages did, where we come from, and where we are going, or why. That is, we don't know what information is relevant, and what information is irrelevant to our lives" [7].

3. WHO DO WE THINK WE ARE?

Although the discussion so far hopes to inspire some sympathy for Postman's position, it is past time now to subject his claims to more robust rational scrutiny. This requires addressing two intertwined puzzles: First, supposing we know what "cultural AIDS" amounts to, what exactly does it mean to say *we* suffer from cultural AIDS? Who exactly is the *we* in question? Second, what sorts of reasons can be given for and against the claim that we suffer from cultural AIDS? What sorts of discourse must we engage in to advance such a thesis?

To a first approximation, "we" is an expression picking out "anyone relevantly like me," where the context of the sentence establishes the standards of relevance. In some contexts, "we" refers to all humans; in others, it might refer to all rational beings as such; in still others, it might refer to all mammals. In the last sentence of the previous paragraph, the meaning of "we" is given tolerably well as "anyone who wishes to advance some thesis of cultural criticism." In any competent and literal use of "we," it cannot happen that the speaker fails to belong to the class of individuals picked out by the relevant categories. If I am an atheist, it would be a mistake for me to say in my own voice and literally mean, "We believe in the Father, who created all that is," but no one is confused when Steve Green says this in [4].

However, when Postman asserts, "We suffer from a kind of cultural AIDS," it does not seem right to interpret that claim as meaning, "Anyone relevantly like Neal Postman suffers from a

kind of cultural AIDS," since the tone and purpose of cultural criticism is undermined insofar as the speaker suffers from whatever debilitation he is attributing to the culture at large. This seems to be a general feature of cultural critique—claims about what ails one's culture will tend to prompt a sort of self-reference paradox. Note, this is not quite the *logical* paradox we encounter in the paradox of the liar, since claims about what culture is doing to "us" admit of degree whereas the logical paradox of self-reference turn on claims that are categorical. However, this *rhetorical* paradox of self-reference is sufficient to create problems for interpreting first person plural assertions in cultural critique.

That this rhetorical self-reference paradox causes a serious rhetorical problem is apparent in light of the second puzzle: What sorts of reasons can be given for and against a critique of one's own culture? Reflection reveals that the rhetoric of such essays substantially relies on the audiences' ability to judge for themselves, and this extends the self-reference problem, since both speaker and audience are members of the culture being criticized. Since the self-reference problem tends to undermine authority, it is striking that such essays typically do not bolster their authority by proceeding in the manner of a scientific paper or a mathematical proof. The reason of this is apparent: the authors are themselves interested parties, and the occasion of their writing is a certain discomfort or disdain with values they see creeping across the wider culture. The claims of cultural critics require qualitative discriminations in terms of what matters and what does not, discriminations that are not captured by the standard tools of the social sciences or deductive logic. Their theses are essentially about *values*. When Carr asks, "Is Google making us stupid?" or Turkle ponders "How computers change the way we think," or Postman asserts "We suffer from a kind of cultural AIDS," their primary purpose is not to coldly describe some process of change. These essays mean to raise the alarm. These essays mean to convince us that something has gone wrong, and they are a call to action to rebuild the culture and remedy what has gone wrong.

These essays extol but do not prominently feature the methods of science. This is as it must be. Perhaps there are dispassionate studies that will show denizens of the information age have more or less cognitive power along this or that dimension, but such studies cannot settle whether or not we are being made *stupid* in the broad and value laden sense Carr plainly has in mind. What statistic could demonstrate that? It is an evaluative claim, and no cold and clinical description will do it justice. Since statistics and empirical studies are not appropriate for the job, the anecdote predominates. After treating the reader to a series of delectable anecdotes, many of which are first person reports of his own increasing "stupidity," Carr is forced to acknowledge that "Anecdotes alone don't prove much." But scientific data can prove even less of the relevant thesis, and whatever there is of it is tacked on as an ornament or at best as some tenuous confirmation of a suspicion originating elsewhere. Carr's admission that anecdotes are not really evidence is followed by just one paragraph about a study contrasting online browsing with deep reading followed by another paragraph that strains to connect that study with work in developmental psychology before we are whisked into an anecdote about Nietzsche's purchase of a typewriter. Scientific evidence makes no further appearance in the essay.

While the desirability of scientific evidence is acknowledged, these essays do not set out to inform readers of such evidence so much as they defend a research agenda that might produce it, and that research agenda is motivated by the anecdotes that animate some moral project. Turkle is fairly explicit about this. After introducing her readers to the problem she sees by way of an anecdote in which a student interprets a Freudian slip as an error in information processing, Turkle explains, "Such encounters turned me to the study of both the instrumental and the subjective sides of the nascent computer culture. As an ethnographer and psychologist, I began to study not only what the computer was doing *for* us, but what is was doing *to* us, including how it was changing the way we see ourselves, our sense of human identity" [10, 1].

The point is not to dismiss these essays for their lack of hard evidence, but to come to terms with the fact that they do not set out to give hard evidence. Any charitable reading will have to acknowledge this. If one scours Postman's book for explicit or implicit premises that, in the manner of a deductive proof, add up to the conclusion, "We suffer from a kind of cultural AIDS," or if one treats the claim as a hypothesis and looks for Postman's refutations of alternative hypotheses in the manner of abductive reasoning, one will be disappointed. Instead of these hallmarks of the sciences, we find the stuff of the humanities: the sharing of anecdotes and introspections, historical exegesis, and the author's way of thinking laid bare to us with an invitation to think along. Postman tells us a *story* in which Western culture has lost its way, so bedazzled by all things science and technology that it no longer knows why anything should matter or even if it should. It is a cautionary tale, but whatever plausibility the reader finds in it will come from something other than hard evidence. I submit that these sorts of essays operate analogously to what Dennett calls "intuition pumps," which are "little stories designed to provoke a heartfelt, table-thumping intuition—'Yes, of course, it has to be so!'—about whatever thesis is being defended…They are the philosophers' version of Aesop's fables, which have been recognized as wonderful thinking tools since before there were philosophers" [2, 11]. The designation "intuition pump" is widely mistaken for a pejorative, as a way to dismiss a certain kind of muddleheaded thinking on the grounds that it is insufficiently logical and rigorous. Dennett has long emphasized that this is not the case, and I have no intention here of disparaging any good intuition pump. If I point to a snake in the grass and declare, "Look! We're in danger!" I offer no argument, but my effort cannot be dismissed as empty rhetoric. To discover if we are in danger, you will have to look for yourself.

To illustrate the point it is helpful to compare and contrast Postman's cultural critique with something from one of my all time favorite thinkers, Ralph Waldo Emerson, writing on behalf of self-reliance:

"If any man consider the present aspects of what is called by distinction *society*, he will see the need of these ethics. The sinew and heart of man seem to be drawn out, and we are become timorous, desponding whimperers. We are afraid of truth, afraid of fortune, afraid of death, and afraid of each other. Our age yields no great and perfect persons. We want men and women who shall renovate life and our social state, but we see that most natures are insolvent, cannot satisfy their own wants, have an ambition out of all proportion to their practical force, and do lean and beg day and night continually. Our housekeeping is mendicant, our arts, our occupations, our marriages, our religion, we have not chosen, but society has chosen for us. We are parlour soldiers. We shun the rugged battle of fate, where strength is born" [3, emphasis original].

In reading Emerson, as in Postman, one does not find premises followed by conclusions or data supporting or refuting hypotheses. Emerson's appeal rather goes something like this: Dear readers, I offer you a description of certain attitudes and affective states that you will recognize in yourself or in your fellow man or both, along with a story of how these might plausibly arise in the environment we share, and this description will resonate with you as an insight whose truth is apparent as if discovered in an observation; if you understand what I'm saying and think it through, you too will *see* that it is so. The passage brings us to see things a certain way without explicitly arguing things are this way. The intended rhetorical outcome is that the reader should see *for himself,* "Yes, of course, it has to be so!"

Note that Emerson's cultural critique threatens to undermine itself in self-reference as much as Postman's. Emerson declares that "we" have become cowards who shun the rugged battle of fate, but this declaration is presented to us as Emerson's own act of courageous self-reliance and in explicit defiance of the expectations of society. If his project succeeds, it will not be true that "we" will be cowards after all. If the bold observations he calls upon the reader to make with him are themselves expressions of courage and self-reliance, it is already false in his mouth to declare that "we" shun the rugged battle of fate, for he is manifestly engaging it. Emerson and Postman both use "we" where it is clear "they" would be more literally correct.

This similarity, however, calls our attention to where Postman and Emerson differ. Although both thinkers eschew piling up evidence and argument for their main verdicts in favor of offering readers inspired narratives that invite assent in the manner of intuition pumps, and although both make claims about the state of their own cultures in the first person plural that raise a specter of self-reference, their background assumptions about the nature and location of authoritative judgments of cultural worth allow these similarities to work in opposite directions. What there is of paradox in Emerson's story only serves to heighten its power to achieve his desired effect, while these same paradoxes threaten to undo Postman's narrative or at least transform it into something less appealing.

Emerson appeals to the honest judgment of the rugged, self-reliant individual as the highest and most sacred authority—if he is right, then your own observation is more trustworthy than any testimony or evidence from him. If you agree with Emerson's assessment that society makes us cowards and that cowardice is a bad thing, then your resolve to stand against society in favor of your own authority is sure to be increased. Alternatively, if you disagree with Emerson's evaluation of society, then you thereby assert your own self-reliant authority while declaring that the influence of society does not make you weak and cowardly. You might or might not agree that society makes *them* cowardly, but you must deny that it makes *you* cowardly. Thus is your resolve to stand upright in favor of your own authority—with or against society—sure to be increased. The intuition pump operates whether you go along with Emerson or resist, since engaging the story itself forces the reader to exercise her own authoritative judgment regarding the influence of society. Thus is self-reliance bootstrapped, a process captured succinctly in the rhetorical question, "Are you gonna let them push you around?" The only psychologically plausible responses are: "No, they're not pushing me around; I'm doing exactly what I want to do," and "Hell no, I won't let them push me around." While not a logical impossibility, it would be exceeding strange for anyone to reply, "Yes."

To be sure, part of the rhetorical appeal of using "we" where one means "them" is to soften the call to a certain brash elitism. The reader is invited to judge that it is not *we* who are cowards but *them*; in making this judgment is found her escape from the undue influence of the wider society. Although there is more than a hint of elitism if the self-reference paradox is resolved by declaring that it is *them* not *us* who are made cowardly by society, that *we* are better than *them* because we enjoy a self-reliance they lack, the reader is not subjected to any external standard of judgment. Her own authority proclaims the superiority of itself and on its own terms. Elitism notwithstanding, there is no room in Emerson's worldview for anything smacking of paternalism or an arrogant disregard for the judgments of others. They are free and even encouraged to respond as they will. What is to be done will have to be done by the reader for herself, and even if she comes to affirm whatever roles society has assigned to her, she will have to do so in the first person. To say, "They're not pushing me around; I'm doing exactly what I want to do," is to judge for one's self who one is and what one most wants to do and become; it is to assert one's self-reliance. While Emerson's call to action is for each individual to do something with her self, the significance of her actions does not end in the first person singular, since her courage is certain to transform the culture at the margin exactly insofar as she sees it in need of transformation; on Emerson's view, the individual is primary and the culture will be whatever it is as a function of all the individuals that make it up. If there is to be an escape from the influence of one's own culture, it will have to come from this direction, from something that is positioned to push back against the culture.

In contrast, Postman's conception of the relation of the individual to the wider culture works in the opposite direction. On his view, the culture is primary and the individual is a function of the wider culture. This is revealed most tellingly in Postman's call to action (echoed in Carr and Turkle and countless others), which asserts "we" must do something about the culture and thereby save "ourselves" from being amused or informed to death. Specifically, we are called upon to affect wholesale changes in the institutions that make us who we are, especially at the level of some common core of education. The key intuition pumped by Postman's essay is the alleged observation that folks today are living shallow and meaningless lives for want of some transcendent ideal we can all get behind, since thinking each for ourselves about transcendent ideals issues in the modern condition of claims and counterclaims that constitutes information glut and inflames cultural AIDS. However, Postman's own authority to judge rightly is undermined by the self-reference of declaring "we" suffer from cultural AIDS, when no one with cultural AIDS would be in a position to say so with any authority. His audience is likewise in no position to have any authoritative judgment, since they too suffer from cultural AIDS and the world is incomprehensible to them. In light of all this, the only charitable way to read Postman's claim that "We suffer from a kind of cultural AIDS, and we need to do something about it" is something like, "*Those other people, not us,* suffer from a kind of cultural AIDS, and *we* need to do something about *them*." The first half of this claim parallels the rhetorical move we observed in Emerson, but the call to action is starkly different, laying bare a paternalistic undercurrent that Postman's view cannot avoid and which calls into question the reliability of whatever intuitions he pumps. Avoiding this paternalism by reducing Postman's call to action to an admitted shot in the dark dispels nearly all its rhetorical force, while conceding this paternalism asserts for some the dubious authority to judge the ways of life of others.

Emerson invites me to ask, "What shall I make of myself?" Postman invites us to ask, "What shall we make of them?" If Emerson is right, I am placed as the ultimate authority regarding the answer to his question. If Postman is right, I am no authority regarding the answer to his question, and those of us who disagree with him will have to take his word and the word of those persuaded by him that they are the true cultural authorities. Since diverse ways of life are so often mutually incomprehensible, Postman's story has set us up to seem to observe cultural AIDS even where each individual knows perfectly well where she stands and who she is.

4. THE ILLUSION OF CULTURAL AIDS

Place Tables/Figures/Images in text as close to the reference as When I ask students whether they are persuaded by "Informing ourselves to death," there is typically a slim majority who find it compelling. When I ask how many of them think they themselves suffer from cultural AIDS, most of those hands go down. Although Postman has succeeded in getting them to see the cultural landscape of the information age as an alien, chaotic, and unnavigable terrain for most of its inhabitants, this observation of how the world must seem to others stands in contrast to their own first person experience. They see themselves as having the resources to discriminate, reduce, and make use of information in their own lives, while they see others as being buffeted this way and that by fads and fallacies, distracted by baubles and spectacles, and ultimately amused to death.

This disconnect between one's assessment of one's own life and one's assessment of how well others are navigating their lives should come as no great surprise, especially if one is willing to entertain the notion that the grounds of meaning and purpose in one's life tend to be deeply personal and contextualized to a degree that renders them opaque to outside examination. Psychologists have studied and documented a range of attribution biases for decades now, so we know even before Postman invites us to look for the character flaws of our fellows that we are prone to see them [5]. In the first person, we appreciate the nuances of the situation and the full range of antecedent inputs to our behaviors, including whatever overarching narratives we tell ourselves to explain the trajectories of our lives. Our observations of others are necessarily less well informed and typically less flattering as a result.

The inability to fully fathom the meaningfulness others find in their diverse ways of life has been widely remarked. Nagel offers several examples in *The View from Nowhere,* his celebrated defense of the significance of the first person point of view across a range of philosophical questions. In his discussion of values he notes, "It is also possible that some idiosyncratic individual grounds of action, or the values of strange communities, will prove objectively inaccessible. To take an example in our midst: people who want to be able to run twenty-six miles without stopping are not exactly irrational, but their reasons can be understood only from the perspective of a value system that some find alien to the point of unintelligibility" [6, 155].

In our better moments, when we strive especially hard to understand others, we might come to appreciate that some or another pursuit or enthusiasm serves as a ground of meaning or organizing principle of their lives, even where that pursuit or enthusiasm leaves us cold. I enjoyed a memorable conversation with a student majoring in accounting whose prodigious philosophical talents prompted me to suggest adding philosophy as a minor or even a double major. I suggested that, although accounting was a perfectly noble and practical career, he might be enriched by a deeper and more extended encounter with the profound. He was grateful of the compliment, but he explained that he was really passionate about becoming an accountant and did not have the time to pursue philosophy beyond the required general education course. He was *passionate* about *accounting* to the point of choosing it over *philosophy*! I was flabbergasted until he went on to explain, in a distinctly cheerful tone, that he had been diagnosed long ago with mild Obsessive Compulsive Disorder, and he enjoyed an intense thrill in his work as all the numbers lined up and fell into place. I had to admit that the study of philosophy would not sit well with such dispositions. I also had to admit that there may very well be those for whom accounting, of all things, might properly serve as one's highest calling.

The education I received from that student is recalled to me every time I suspect one of my fellows of leading a shallow life of quiet desperation, and it leads me to question the wisdom of flippantly dismissing the ways of life of others without the least effort to appreciate how their ways of life make sense to them from the inside. In contrasting a conception of happiness as mere shallow pleasure with her own more lofty conception involving real accomplishment, "the measure of his success in the service of his life," Ayn Rand disparages those who pursue "'mindless kicks'—like the driver of a hotrod car" [9, 31]. Alas, had she immersed herself in the actual culture of hotrodders, Rand might have found in them an epitome of the happiness she extolls. Had she bothered to see things as the exemplar of the hotrod culture sees them, she would have experienced a profound aesthetic appreciation for well tuned and carefully crafted machines, along with a profound respect for the artists responsible for creating them and keeping their myriad of parts dancing together in well lubricated harmony. The sublime rush of the hotrod roaring to life around one, singing of its muscular power and grace, pounding its chest in steady rhythms conveyed to the breast of the enthusiast and shaking the very earth all around, all this is encountered as noise and nuisance to the uninitiated. Hang around the exemplars of hotrod culture long enough, and you begin to get a glimmer of what their enthusiasm is all about. They are not generally attuned to subtle discussions of Aristotle or Nietzsche, but they have their own rich texts and interpretations of those texts, their own heroes and idols, gods and altars. Of course, the experience of the exemplar of hotrodding is liable to be as far removed from that of the typical hotrodder as the experience of any virtuoso is removed from the less talented or the novice, but it would be a mistake to dismiss even the least sophisticated among them as wasting their shallow lives on pursuits with nothing more to offer than "mindless kicks." It is one thing to observe that a man is not very far along his chosen path in life; it is quite another to suppose he is on no path at all. There is good reason to suspect that the latter is more often an illusion than an observation.

5. PATH DEPENDENCE IN CULTURE

The various paths of our fellow travelers are routinely invisible to us, and this invisibility can prompt us to suppose they are on no path at all. If we project this supposition into our discussions of culture in the information age, it appears to us in an illusion of cultural AIDS. One's interpretation of her own life and her interpretations of others' lives are equally dependent on the path she finds herself on, and the distortion in her field of view caused by the magnetism of her own path is amplified by another kind of path dependence, the sort we observe in the concrete history of anything which evolves. Whether one finds the hotrodder's way of life compelling or alien to the point of unintelligibility will

depend on one's past experiences and the lessons learned from these. This will be so even if we aspire to the utmost objectivity and reason we can muster. It is often perfectly rational, in an internalist sense, to believe one thing or another depending on the order in which one hears the arguments. Whether the argument from evil is compelling may depend on whether one hears it before or after one hears Kierkagaard.

Given the arbitrary and accidental nature of any concrete history, this means we inevitably find ourselves on different paths, as matters of no deep significance in themselves take on a profound ability to set the very horizons of our thoughts and serve as the grounds of our values. Any step out of place along the way, and individuals can wind up in vastly different places. Ironically, one mechanism of path dependence is precisely the filtering and reducing of information Postman attributes to a healthy cultural tradition. That is, path dependence is especially strong for those who do not suffer from culture AIDS, and the diversity of perspectives that issues from this path dependence is an important factor in creating the illusion of cultural AIDS. This path dependence of individuals and of the various and overlapping dynamic traditions that inform their lives ensures a degree of permanent and reasonable pluralism that stands in tension with the calls to action proposed by Postman, Carr, Turkle, and other would-be improvers of the culture.

A cursory examination of internet culture confirms that we live in a time of robust debate, playful repartee, and earnest multilogues that cannot be easily explained if we live in a time of widespread cultural AIDS. Our culture in the information age is a seething, churning, bubbling cauldron of ways of life, wisdom traditions, transcendent ideals, and cat videos. And besides these are also pornography and monster trucks and rock-n-roll and psychedelia and rude jokes and dancing hamsters and fan fiction and hipsters photographing their food and others photographing hipsters photographing their food, and so much more. Postman worries that we are drowning in all this information, and he intimates that we need to do something about it, that we need to rebuild the culture, that we need to make the world safe again for transcendent ideals. In the end, he is contending that we need to save all these pour lost souls whose lives are so empty and shallow without some steady hand at the rudder—ultimately, as we have seen, this has to mean *his* steady hand along with the hands of those who see things *his* way.

I, for one, do not see things his way. I see a diversity individuals gathering and disbanding in a diversity of swarms and flocks and herds and schools and teams and such, each overlapping one another in a myriad of ways and along a myriad of dimensions, sometimes in contest with one another, sometimes in concert, each in various degrees inscrutable to outsiders but each rendering the world intelligible and navigable by their own lights, if only for a fleeting moment or as an ill-considered hypothesis. I contend that this is enough. As we stake our many claims online, it becomes clear where we stand (or shuffle or dance), and this clarity intimates how we make and remake the culture every day, without any navigator, shepherd, or philosopher king.

The illusion of cultural AIDS rests upon some observation like, "If *I* was a person wasting my days on cat videos, pornography, and monster trucks, then *my* life would be tragically shallow and devoid of meaning." I see no reason to dispute such observations, but such observations do not entail: "That person's life is tragically shallow and devoid of meaning." There is nothing objectionable in prodding such a person to find out whether she really thinks her life is all she might hope for (it probably isn't) and whether she wouldn't be better off taking on some more profound and transcendent ideal to organize her life and make sense of the world, but if she resolutely declares contentment with her life as it is, especially after she has been exposed to the testimony of others and the many alternative options available to her, then the testimony of others is moot. We shall have to take her word for it, for she is the only one situated to judge. For all that, she could be wrong; we might hope she would at least try on some other ways of life and seriously consider her alternatives. But at the end of the day we shall have to concede final authority to her. This concession expresses our epistemic humility; though she may be wrong, we cannot know this *a priori.* Whether she is wrong or we are wrong will have to be determined in the course of her experiments in living. Only thus accomplished will this determination be to our mutual edification.

This affirms Emerson's insight that each individual is best situated to operate on her own authority with respect to the true, the beautiful, and the good, and no one is situated to operate as an authority over the culture as a whole, for any such authority would have to issue decrees concerning matters he cannot properly understand, matters over which *she* is the best authority and not any *we*. This deep fact is obscured when we think in terms of what the internet is doing to "us" and what "we" should be doing about it, succumbing to the fatal conceit that someone already knows the outcome we hope to discover and that it is easy or even possible to design and implement some better culture. The alternative I propose is me thinking for myself in terms of what I should be doing and you thinking for yourself about what you should do and so on for each, letting the wider culture emerge as a function of my experiments alongside and in conversation with yours and those of so many others, each in our own spaces and according to our own concerns as these emerge from our own concrete histories, arbitrary and path dependent as these inevitably may be. What grounds meaning for each individual is liable to be incomprehensible across individuals, so there is liable to be no shared conception of the good life operating across the culture at large, and this may create the illusion of cultural AIDS and misplaced worries that "we" need to turn our utmost attention to discovering or creating some single conception of the good that "we" can all rally around, lest "we" suffer from decadence and escapist distraction, amusing and informing ourselves to death, when in fact the absence of any culturally shared conception of the good life or any transcendent ideal before which all are expected to bend the knee opens up a space wherein individuals are able to navigate each according to her own north star. Such considerations suggest that a culture that takes no stand on the meaning of life may leave the most room for individuals to find it and enjoy it for themselves on their own various terms, accommodating the fact that these terms are sure to be especially various in the information age.

With millions and even billions of individuals muddling their ways through the grounds of meaning, each informed by a million particular reasons of their own and the path dependencies these entail, each moving by small steps from worse to better conceptions of the good by their own lights, it is the height of arrogance to suppose anyone knows how to design a better culture—one that will better satisfy all these diverse individuals' hunger for meaning and purpose. If we are to agree with Postman and Carr and Turkle that "we need to rebuild the culture around information technology" [10, 4], let us take care to read that "we" as "each of us, by our own lights, as fierce individuals though we toil side by side and in conversation with one another."

The difference between my view and the alternative presumptions of Postman and so many other critics of the culture is not to be found in the significance we ascribe to culture in establishing the grounds of meaning but in the way we conceive the flourishing of a culture in its fulfillment of that role. Our efforts are equally directed to discovering the best arrangements for the farming of the cultural landscape; we each mean to discover the greatest bounty this soil will support. Our difference lies in our instincts regarding the organization and regulation of those who will till this land and what resources they shall have at their disposal. Some intuit against all empirical evidence that central planning and authority will ensure the greatest yields, the most beautiful vegetation, and the most nutritious produce. They invite us to imagine that every farmer should be a peasant in thrall to some single lord and master, lest the vacant throne of some transcendent authority declaring its single vision of what must be done should spell an anarchy of fields lying fallow in the full bloom of early summer as erstwhile laborers rest lazily beneath shady trees or wander, aimlessly grazing on whatever wild cereals should sprout unattended by human hand. Against this cultural feudalism stands a vision of a diverse multitude of yeoman farmers, each applying his own rugged industry to his own corner of land, each engaged in a series of bold experiments and trials, each attuning his understanding and techniques to the particular and local features of the land and its full potential. That there will be among them loafers and grazers it may be safely admitted, but it is certain beyond hope that this will become a way of life for them only so far as it produces what they need and no farther. Observing the many trials all around him, each will discover for himself what sorts of expenditure and in what degrees yield the greatest returns in the fullness of the season, and he will learn in course to make his own estate as much of a garden as fulfills his own vision of paradise. Instead of being yoked and bent to some fragile and arbitrarily imposed monoculture, exposed to whatever pestilence might sweep across the land to burrow into the roots of his assigned livelihood, he finds himself upright and flexible, prepared to adapt as necessary to keep his homestead and its environs in beautiful bloom. .

6. REFERENCES

[1] Carr, N. 2008. Is Google making us stupid? *The Atlantic* (July/August 2008), http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/.

[2] Dennett, D. 2013. *Intuition Pumps and Other Tools for Thinking*. W.W.Norton & Company, New York.

[3] Emerson, R. 1841. Self-reliance. In *Essays: First Series*. http://www.emersoncentral.com/selfreliance.htm.

[4] Green, S. 1991. We believe. http://stevegreenministries.org/product/we-believe-7/.

[5] McRaney, D. 2012. *You Are Not So Smart*. Oneworld Publications, Oxford.

[6] Nagel, T. 1986.*The View from Nowhere*. Oxford University Press, Oxford.

[7] Postman, N. 1990. Informing ourselves to death. https://w2.eff.org/Net_culture/Criticisms/informing_ourselves_to_death.paper.

[8] Postman, N. 1992. *Technopoly*. Random House, New York.

[9] Rand, A. 1961. *The Virtue of Selfishness*. Penguin Books, New York.

[10] Turkle, S. 2004. How computers change the way we think. *The Chronicle of Higher Education*. 50:21, B26.

.

Machine learning in decisional process. A philosophical perspective

Teresa Scantamburlo
DAIS, Università Ca' Fosari
via Torino 155
Venezia, Italy
teresa.scantamburlo@unive.it

ABSTRACT

In this paper we would like to undertake a critical examination of machine learning in the context of data revolution. Starting from the existing literature, which has in fact highlighted potential risks both at the epistemological and ethical level, we will try to suggest the main limitations of an intensive application of machine learning to decision making. Our discussion we will make direct reference to Satosi Watanabe, whose contribution springs from a genuine reflection on machine learning research and rises many philosophical questions. In addition, we will consider the difficulties of machine learning by exploiting the classical distinction between "apprehension" and "judgement" recalled even more recently by some studies dealing with the emergence of complexity in human cognition. Rather than being an exhaustive analysis, our investigation is a tentative step towards a better understanding of machine learning and its potential implications on individual and social life. Our main contribution is to try to introduce in the philosophical debate some new considerations which come from the inner core of machine learning and from the traditional notions of philosophical logic.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues—*Ethics*

General Terms

Human Factors, Design, Algorithms

Keywords

Machine learning, big data, decision making, judgement

1. INTRODUCTION

Many philosophers and social scientists did not remain indifferent to the profound changes produced by the new information and communication technologies. In the recent years, for example, several efforts have been devoted to the issues regarding the explosion of data availability and its subsequent elaboration by means of apposite algorithms. We commonly refer to this phenomenon by the term "big data" and many a scholar have already highlighted several ethical, as well as epistemological, aspects (see, e.g., [9, 13, 14]). In this paper we would like to carry on such a discussion paying particular attention to the ethical dimension. Specifically we would like to consider the main concern for the pervasive role of data mining processes to assist human activities. Indeed, ranging from advertising to health care and policing we are witnessing an ever-growing demand for access to machine learning research (e.g.: pattern recognition and data mining) [1] to the point that it seems that we are "on the cusp of using machine learning for rendering basically all kinds of consequential decisions about human beings" [12].

As some studies suggested [4], the insensitive application of machine learning could bring about unexpected societal harms especially to racial minorities and disadvantaged people. This runs counter to a commonsensical view: the assumption that algorithmic decisions are "fair by default" [12]. A belief that is often accepted because we think of uncritically the relationship between algorithms and accuracy, as if the mathematical and statistical aspects of algorithms were a certain sign of reliability. But, contrary to what we used to think, machine learning procedures might give a strong incentive to discriminatory decisions even in the absence of discriminatory purposes. This may occur, for instance, when a machine learning function generalizes the sample of a biased population, where some groups are under- or over-represented. Its results will be likely to reproduce the prejudices encoded in the training data, i.e., the data that are used to "teach" the machine learning system to behave in a certain way. The outcome could be simply traced back to the analyst's biases, kept hidden by the "opaque" environment of algorithms, without going further into the problem. But, as Solon Barocas and Andrew Selbst put it "discrimination may be an artefact of the data mining process itself, rather than a result of programmers assigning certain factors inappropriate weight"[4, p. 3]

Our aim is to deepen the discussion about the intrinsic limi-

[1]Note that although throughout the paper we will use the term "machine learning," much of our discussion could be referred to pattern recognition and data mining as well. Indeed, despite the various characterizations we would like to consider all these camps as part of the same endeavour, that is, the attempt to model inductive and generalization processes

tations of machine learning procedures also from a non technical angle and specifically by addressing the philosophical meaning of machine learning. In so doing we will suggest further motivations to the critical analysis of the massive application of machine learning in decision-making. Specifically our investigation will make direct reference to Satosi Watanabe, who is one of the pillars of pattern recognition and machine learning research areas. As we will see, his contribution, which springs from a genuine reflection on machine learning problems, is full of philosophical insights and, we think, represents a solid foundation for the critical analysis of machine learning research. Moreover we will approach the philosophical difficulties which arise when we adapt the notion of judgement to algorithmic learning. Starting from the classical distinction between "apprehension" and "judgement" and exploiting further psychological conjectures, we will argue that a machine learning system can model "only" the early stages of cognition, i.e. apprehension, which is in principle the act by which humans form concepts and, more in general, animals respond to sensorial stimuli. But the formulation of a judgement, which requires consciousness and the ability to compare two or more apprehensions acquired at different times, is beyond the reach of algorithmic decisions. In conclusion our overall discussion will try to suggest that the inadequacy of machine learning in dealing with everyday human decisions can emerge from a number of technical constraints which may reflect, if not exasperate, existing social inequalities. But such a difficulty can be more profoundly understood through the lens of philosophy and the philosophical interpretation of machine learning results.

2. MACHINE LEARNING AS A DATA DRIVEN SCIENCE

Machine learning is an integral part of artificial intelligence. Basically, it deals with the computational processes that underlie learning in both humans and machines (for a general introduction see, e.g., [15]). From a technical point of view its main purpose has been commonly associated to the problem of induction and to the design of algorithms which are able to generalize from a given set of examples. The idea behind a machine learning algorithm is to observe a certain number of objects or events and find a (decision) rule which classify the future instances making as few errors as possible. Note that in the machine learning literature there exists a common distinction between supervised learning or classification, which is conducted without training the system with specific examples, and unsupervised learning or clustering, which is produced, on the contrary, without any external cues [2].

A basic formulation of a machine learning problem could be the following. We define two kind of spaces: the input space $X = \{x_1, x_2, ..., x_n\}$ (also called space of instances) and the output space $Y = \{y_1, y_2, ..., y_n\}$ (known as the label space). For example, if the task is to classify two sorts of fishes (e.g., sea bass and salmon), the set of instances could be represented by a list of characterizing features (e.g., colour, weight, high) into a vector space, and the set of label could be defined by a binary variable, e.g. $Y = \{+1, -1\}$ where the first value stands for "the class of sea basses" and the second one for "the class of salmons". The main goal of classification is to find a classification rule (the classifier) $f : X \to Y$ which map the space of the instances into the label space. In order to learn such a function, the algorithm is presented with some examples (the so-called training data) which consists in a set of pairs of instances with their corresponding labels, $\{(x_1, y_1); (x_2, y_2); ...; (x_n, y_n)\}$. If the training set systematically presents a group of features with a defined label, the algorithm will learn to assign that label to the instances with the same features. Nowadays this strategy is used to perform a variety of classification tasks, with different degree of complexity, such as spam detection, face or speech recognition, customer segmentation and medical diagnosis.

Like artificial intelligence, machine learning has developed several approaches over the years (for a recent review see, e.g., [18]). A first one traces back at the early days of the field and relies on the famous Newell's and Simon's physical symbol system hypothesis. According to the symbolic view every intelligent behaviour, including learning, could be expressed as a manipulation of symbols which encoded specific domain knowledge. Following Kuhn's view of science, a recent study [8] claimed that this approach represented a real paradigm (the knowledge-driven paradigm) in the development of artificial intelligence. But most importantly, this study acknowledged that artificial intelligence - but the same holds even for machine learning - has traversed a paradigm shift passing from a knowledge-driven to a data-driven approach. Contrary to the previous paradigm, the current one does not aim at designing general cognitive skills or expert systems, but it tries, more modestly, to reproduce the input-output behaviour available "in the wild". In this paradigm shift, the growth of data resources represented a determining factor which has driven the development of many successful applications ranging from statistical machine translation to computer vision (for a manifesto of the data driven approach to machine learning see [11]).

Now, the data-driven paradigm is often celebrated as a triumph of big data and as a phenomenon which goes far beyond the boundaries of machine learning and artificial intelligence. According to Jim Gray, for instance, this paradigm shift involved the entire practice of science, including traditional research areas like the humanities and the social sciences. As a consequence, the data-intensive science, what Gray called the "fourth paradigm" [21], encourages not only a greater respect for data collection and conservation, but also a wider application of machine learning. The latter, in this way, started getting involved in a number of tasks which were considered to be the traditional work of culture [3]: advising people or places, searching meaning, sorting relations, etc. Probably, the digital humanities and the computational social sciences offer the most exciting applications.

However, despite the general excitement and the various beneficial results, the data-intensive research prepared the

[2]From now on, when we will speak about machine learning we will refer exclusively to the problem of classification

[3]A similar discussion has been done by Ted Striphas about the role of algorithms in culture and the phenomenon of algorithmic culture. For more details see an interview available at: https://medium.com/futurists-views/algorithmic-culture-culture-now-has-two-audiences-people-and-machines-2bdaa404f643

ground for a widespread ideology, based on the idea that big data and new data analytics could give us "a full resolutions on the worldwide affairs"[19]. According to this new form of positivism the data-intensive research can supersede the traditional ways of knowing and, most importantly, encourages new business circles[4]. One of the most cited supporters of this new approach is Chris Anderson who wrote a famous piece on the outstanding advantages of data revolution: "This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves." [1] Clearly this position arose a fiery debate within research communities and a wave of critical reactions pointed out both epistemological and ethical concerns. Ultimately, the contemporary debate is making clear that there is a urgent need for a critical reflection on the role of machine learning within society, a task that could be favoured by a philosophical discussion of the meaning of machine learning and its ethical impact.

2.1 Machine learning and its ethical impact

The problems related to the massive application of machine learning are drawing attention both in private and public institutions. For example, President Obama called on the administration to conduct a 90-day review of big data and privacy. This survey resulted in a report titled "Big Data: Seizing Opportunities, Preserving Values", which arose problematic issues about the role of new data analytics in individual and social life. One of the main findings of the report is concerned with the "opaque" environment of machine learning which could bring about unintended discriminations within society. Specifically, the report pointed out that "some of the most profound challenges revealed during this review concern how big data analytics may lead to disparate inequitable treatment, particularly of disadvantaged groups, or create such an opaque decision-making environment that individual autonomy is lost in an impenetrable set of algorithms" [16, p. 10]. Interestingly the same concern started attracting the activity of research institutions and scientific communities, included the area of machine learning where specific initiatives on the theme have gained considerable resonance throughout the community[5].

Starting from the provoking issues of the aforementioned report Salon Barocas and Andrew Selbest tried to shed light on the "impenetrable" environment of machine learning so as to suggest which steps could be more critical for decision-making. Scrutinizing the main phases of machine learning procedures they suggest that each step could create the "possibilities for a final result that has a disproportionately adverse impact on protected classes, [...], failing to recognize or address statistical biases, reproducing past prejudice, or considering an insufficiently rich set of factors." [4, p. 5] This conclusion collides with the religious adhesion to the data-driven science and contributes to create a more critical approach to machine learning and to the ways in which this field can be applied. But let us consider some specific difficulties.

A first problematic aspect of machine learning is the definition of the class label, also known as the target variable. In the example of the previous subsection, defining the "class of salmons" could be a straightforward task because the identification of some distinctive features could be enough to recognize a member of that category. This may occur frequently if we have properties which are good in their diagnostic function, but this is not always the case and reality could be more complex. Indeed, as Barcoas and Selbest pointed out, there are many other classes which are hard to define like the case of "creditworthiness" or the definition of a "good employee". For example, consider the second example: what are the cues for a good employee? Is it the amount of time he/she spends in his/her workstation or his/her production time? The number of tasks he/she is able to fulfil in a month? "These may seem like eminently reasonable things for employers to want to predict, but they are, by necessity, only part of an array of possible ways of defining what "good" means. An employer may attempt to define the target variable in a more holistic way – by, for example, relying on the grades that prior employees have received in annual reviews, which are supposed to reflect an overall assessment of performance." [4, p. 9]

Actually, the problem of defining the target category is inherently correlated to the intrinsic limitation of essentialism which has had a profound influence on the field of machine learning, as well as on other disciplines (for more details on the essentialist paradigm in the machine learning area see [17]). Indeed, the evolution of the field is tied to the notion of essential features so that the whole field can naturally be cast as the problem of finding the essential properties of a category. However, even though the feature based approach presents a number of advantages because of its geometrical interpretation – features indeed can be mapped into a features vector space –, there are numerous application domains where either it is not possible to find satisfactory features or they are inefficient for learning purposes. These difficulties are bound to increase if we consider the ever-growing diffusion of machine learning in domains as diverse as education, economy, policy, etc. The more we will use machine learning and the more we will get involved in the problem of finding appropriate features and definitions which are able to deal with the complexity of human life.

In a way big data could give us the illusion that extending the range of examples the classifier could learn better and more about the nature of a category. The idea is simple: if we have a potentially infinite number of instances to train the algorithm, we will be more inclined to think that the extracted features convey the very image of the class that we want to represent. However, even the idea of extending the training set shows further difficulties that should not be ignored. Indeed, when the training phase includes the data released by users' interactions (e.g., the tweets or the messages exchanged in a social network) it may reproduce the preju-

[4]On the role of big data within economics see, e.g., [22].

[5]For example a specific discussion on the topic was conducted during a workshop titled "Fairness, Accountability and Transparency in Machine Learning" organized in some of the leading conferences of the field, i.e. NIPS 2014 and ICML 2015. More information are available at `http://www.fatml.org/`

dice of individuals. In this way, the biases encoded in user's actions are involuntarily passed to the algorithm which take those actions (e.g, single click or typed words) as input and, thereby, as reference to learn a decision-making rule. An interesting example of the concrete implication of this difficulty has been given by Latanya Sweeney who discovered "that Google queries for black-sounding names were more likely to return contextual (i.e., key-word triggered) advertisements for arrest records than those for white-sounding names." [4, p. 12] Sweeney's study suggests that such a result is not produced by some external factors (e.g. the interest of advertisement companies) but by the algorithmic procedure which Google uses to select the advertisements to present alongside the result of certain queries. Thus, the advertisements appeared on the basis of the historical trend of user's clicks: "an advertiser may give multiple templates for the same searching string and "Google algorithm" learns over time which ad text gets the most clicks from the viewers of the ad. [...] At first all possible copies are weighted the same [...] Over time, as people tend to click one version of ad text over others, the weights change, so the ad text getting the most clicks eventually displays more frequently." [20, p. 34]

Moreover the idea that users' behaviour could feed machine learning algorithms suffers from another important limitation. This is concerned with the process of data collection and the concrete possibility that groups of people, especially those living at the margins of the big data landscape, are systematically neglected. In these respect, even though one of the promise of big data is the increase of people inclusion and participation, "there still exists dark zones or shadows where citizens and communities are overlooked or underrepresented." [7] This may occur, for example, when people of protected classes are prevented from providing feedbacks about a specific service because of the differences in the Internet access. In this case the algorithm which has to reveal some information about that service will reflect only a part of users' opinion and its predictions will be not so reliable.

Finally, the outline of some potential risks underlying the machine learning process suggests that the data-intensive approach requires a mindful application and an in-depth analysis of the human factors involved throughout the whole implementation of the system.

2.2 Ambiguity in machine learning

So far we have seen that there are several insidious steps in the overall machine learning procedure. This resulted from a detailed analysis of the potential discriminatory impact of machine learning. But, a similar conclusion could be naturally drawn from the intrinsic ambiguity of inductive inference which is the process that machine learning tries to mechanize. One of the early accounts of this ambiguity has been given by Satosi Watanabe [23], a father of pattern recognition area. In his seminal book, *Pattern recognition: Human and Mechanical*, Watanabe argues that the inductive process involves a basic indeterminacy that cannot be removed by experimental data or logical dependencies.

On the arbitrariness of inductive inference we know that a first eminent contribution came from David Hume, who holds, notoriously, that induction has no logical foundation. But in spite of Hume's lesson, Watanabe pointed out that there exists a persistent tendency among philosophers and mathematicians to reduce induction to logical inference. According to Watanabe, this widespread bias stems from the "necessary view" of probability and hinders a correct appreciation of the depth and breadth of inductive ambiguity. The most systematic account of this logical interpretation was given by Rudolf Carnap[5]. In his formulation we find that the probability $c(h, e)$ of an hypothesis h, given the evidence e, is strictly depending on the relation between h and e and any other elements outside h and e (e.g., human opinion) can play a role.

More or less indirectly this view is still present in our society and the ascent of data driven science gave surely new incentives to it. Indeed, the advocates of data driven science leverage precisely the power of logical inference, which is intended by definition as a source of objective knowledge, and the abundance of data distributed over the net. An interesting example of this "faith" is give by Ian Steadman who described how we will typically gain knowledge in the near future: "The algorithms find the patterns and the hypothesis follows from the data. The analyst doesn't even have to bother proposing a hypothesis any more. Her role switches from proactive to reactive, with the algorithms doing the contextual work"[19]. Thus, when we adhere to such view, human intervention appears almost meaningless. This is the idea, for instance, that has been expressed in reference to the launch of Ayasdi, a data visualisation software which uses big data: "by using algebraic topology Ayasdi has managed to totally remove the human element that goes into data mining – and, as such, all the human bias that goes with it. Instead of waiting to be asked a question or be directed to specific existing data links, the system will – undirected – deliver patterns a human controller might not have thought to look for"[6].

But is this view really tenable in the case of induction? Watanabe answered clearly "no". The reason is that, as well as h and e, the probability $c(h, e)$ includes further extra-evidential factors which are extraneous to the "necessarily relation" between the hypothesis and the evidence, and are ultimately determined by human choice. When we consider induction in its Bayesian formulation most of the extra-evidential factors are included in the prior probability of a certain hypothesis whose credibility is not determined only by experimental data but also by the original extra-evidential evaluation of it. Among extra-logical factors that induction could involve, Watanabe considers a variety of aspects which are analogous to those operating in scientific discovery. For instance, Watanabe cites the criterion of simplicity ("scientists often believe in one hypothesis more than another because it is simpler or more elegant"[23, p. 104]), or the overall coherence of a theory ("Two hypotheses which are equally well confirmed by the evidence may have different credibilities depending on how well they harmonized with other hypotheses in a theory"[23, p. 104]).

Exposing the values and the general considerations which can operate, with logical relations and evidences, in the inductive process, Watanabe made an implicit reference to the complexity of scientific activity and its analogy to machine learning inference. Machine learning, indeed, has always

looked at the process of scientific discovery as a fundamental model for its development and, hence, as a source of inspiration. But now, the application of machine learning to other aspects of intellectual and moral activity introduces further complexity to the design of appropriate solutions. This leads us to think that the employment of machine learning in a increasingly broad range of experiences will surely expand the list of extra-evidential factors affecting machine learning inference.

3. HUMAN JUDGEMENT AND ALGORITHMIC DECISIONS

As we have seen before, the final goal of machine learning is to obtain a decision function. But in order to find such a function the system need to learn. But what does it mean to learn in machine learning? Recalling Watanabe's contribution we might say that, given a repetitive environment, "learning is a process in which the response behaviour converges to one of the alternative possibilities"[23, p. 115]. At the beginning of the process, the response behaviour appears disordered and any pattern seems to emerge but as the learning progresses the behaviour shows a clearer direction until it focuses on fewer and fewer alternatives. From a psychological point of view this dynamics corresponds to the process of pattern formation (what Watanabe called "conceptual morphogenesis"), which starts from a condition of unstructured data, a uniform distribution of points, and through repeated mental adjustments, it comes to see a "well-structured" or a "well-organized" set of data (i.e., a pattern or a form). In this way, the learning process can be defined as an entropy-decreasing function whose main objective is to achieve small values of entropy.

This definition, however, could be presented in a slightly different way. In many cases, indeed, learning is characterized as a convergence towards the "right" response alternative. But, as Watanabe suggested, "learning a wrong alternative is also a learning" [23, p. 115] as the evaluation of the correctness of a response behaviour is a step beyond learning. Therefore, according to Watanabe we should clearly keep distinguished the evaluative aspects from the descriptive aspects of learning. It is precisely from this distinction that we would like to deepen the role of machine learning in decisional process. Indeed, Watanabe's observation invites us, more or less indirectly, to understand machine learning as a process which is not immediately concerned with judgement. Rather, its role is much more similar to the notion of apprehension, one of the key concept of classical philosophy, which is substantially different from that of judgement. A brief discussion of these two cognitive skills will allows us to better understand the meaning of machine learning procedures and its intrinsic limitations.

3.1 The nature of simple apprehension and judgement

In the history of philosophy the distinction between apprehension and judgement is a very classical one. It traces back to Aristotle and in philosophical logic recalls the tripartite division of the acts of mind (i.e., apprehension, judgement and reasoning). In a nutshell, when the mind is neither affirming nor denying the act is called apprehension, when it is making an affirmation or a denial the act is called judgement and, when this decision is mediate, i.e., the conclusions are drawn from previous judgements, the act is called reasoning. From a psychological point of view they expressed different stages of cognition, where the upper layers make use of the material provided by the bottom ones. Now, we will briefly outlined the first two stages (i.e., apprehension and judgement) without going into much detail.

Simple apprehension is the starting point of knowledge processes. It is the act of perceiving an object intellectually, without affirming or denying anything about it. It correspond to the intellectual grasping of the object, what results from the process of abstraction, which starts from some perceptual stimuli and end up with a conceptual form. The adjective "simple" tends to emphasize the fact that an apprehension affirms nothing and denies nothing, it simply conceives the idea of some object. On the contrary, a judgement is an act of the mind which joins or separates two terms through affirmation or negation. Unlike apprehension, which is a form of direct understanding, a judgement is a reflective act of mind which is able to combine more conceptual terms (i.e., more apprehensions) and to compare such a synthesis with the external experience. It is only with judgment that the mind fully commits itself to a truth or to a falsehood.

The distinction between apprehension and judgement has been recently recalled by Tito Arecchi to describe two different aspects of decision-making (e.g., see [2, 3]). In his formulation apprehension is characterized as a coherent perception which emerges from the recruitment of neuronal groups, while judgement is presented as the ability to compare two or more apprehensions coded in a suitable language. Specifically, apprehension is presented as a dynamical strategy which lasts around 1 sec and is a byproduct of a collective synchronization of different neuronal areas, each of one behaving as a chaotic dynamical system. It results in a motor reaction and is common to all higher animals. More formally, apprehension can be expressed as a Bayesian inference, $P(h^*) = P(h|d) = (P(d|h)P(h))/P(d)$, where h^* denotes the most plausible interpretative hypothesis which is selected in presence of a sensorial stimulus d, and $P(d|h)$ is the procedural model which represents the equipment whereby a cognitive agent faces the world. This type of inference takes place when a bottom-up stimuli arrives and the top-down elaboration exploits a set of models $P(d|h)$ retrieved from the memory selecting the hypothesis that best fits the actual stimuli.

On the contrary, the process of judgement requires the comparison of two apprehensions acquired at different times, coded in the same language and recalled by the memory. It lasts around 3 sec and requires "self-consciousness, since the agent who performs the comparison must be aware that the two non simultaneous apprehensions are submitted to his/her scrutiny in order to extract a mutual relation." [3, p. 323] By analogy with the Bayesian formulation of apprehension we call d the code of the second apprehension and h^* the code of the first one, which is now already given. Now, as opposed to apprehension, in the process of judgment the cognitive agent is asked to retrieve $P(d|h)$, which represents the conformity between d and h^*, that is, the best interpretation of d in the light of h^*. In this case, there

is no algorithm presupposed and the agent has to build a new one through an inverse Bayes procedure. An example could better suggest the difference: a rabbit perceives a rustle behind a hedge and it runs away, without investigating whether it was a fox or just a blow of wind (apprehension), whereas, to catch the meaning of the fourth verse of a poem, we must recover the third verse of that same poem, since we do not have a-priori algorithms to provide a satisfactory answer (judgement).

Ultimately, the difference between apprehension and judgement can be interpreted in the light of Gödel's first incompleteness theorem and Turing's halting problem. The idea is that while apprehension is a deterministic process, formally described by an algorithm, judgement is a creative act which implies a holistic comprehension of the surrounding world (semiosis). From a mathematical point of view this coincides with the difference between climbing up a single slope of by a steepest gradient program and jumping from a slope to another one in a multi-mountain landscape. The ability of judging is what discriminates human living from Turing machines and is the source of creativity and decisional freedom.

4. CONCLUSIONS

Turning back to the field of machine learning we could draw some conclusions. In the first place, we could better understand the meaning of machine learning procedures. They can render the process of apprehension, which measures the fitness of a perceptual stimulus to a given interpretative model, but they cannot formalize e real judgment, which requires, on the contrary, consciousness and, hence, the ability to reflect on acquired apprehensions. This means that machine learning results are not self-evident in view of the fact they do not imply any real commitment to a truth or to a falsehood. Contrary to the supposed idea that machine learning results, like numbers, can speak for themselves, we think that there are intrinsic motivations in favour of a in-depth evaluation of algorithmic outcomes, a task which requires the creative work of judgment and reflection. As Watanabe suggested, the judgement as to the "right" and the "wrong" behaviour is an activity which comes after apprehension and which needs a very human intervention, especially in those applications which have a direct impact on social life (e.g., think of machine learning procedures applied to employment).

Human intervention, as we have seen, is bound to became greater and greater. The contribution of humans in data mining process is intrinsically motivated by the ambiguity of inductive inference which is based on several evaluations (e.g., prior knowledge) as well as data and logical relations. In other words, this says that a certain degree of interpretation is always included in the mechanization of induction. Unfortunately, when this interpretative work is extended to a broad range of people, who are (implicitly) engaged in the data mining process, it becomes more complex to model a collective decision. An indirect sign of this complexity could be found in the disparate effects of data driven science and the challenge of producing automatically fair decisions. On the other hand, that fair decisions can be produced by running a machine learning algorithm on large pools of data is an overly simplistic idea that ultimately assumes that the human decision is fully described by the collective synchronization of different neuronal areas.

In the end, our arguments provide fresh support to the criticisms raised against the myth of objectivity of data-driven science. As suggested in [9], Anderson's dismissal of all other theories and disciplines is a naive thought which does not seriously consider the rationale of machine learning work. Raising machine learning to the rank of higher form of knowledge is a dangerous move which formally reduces perceptual knowledge (i.e., apprehension) to the reflective activity of judgment. From an ethical point of view this operation is somehow analogous to the one assimilating moral judgment to a simple perceptual model, where moral decisions are the result of a "quick, automatic evaluations (intuitions)." [10, p. 814] Often such models arise as a reaction to the overestimation of reasoning in the study of moral action. As a consequence they tend to view morality not as a process of ratiocination and reflection but rather as a process more akin to perception. However we suspect that the study of moral behaviour cannot be reduced to the rationalist-intuitionist debate, as to say that moral action is either reasoning or perception, but requires a more attentive analysis of the continuum and the complexity of human cognition.

5. REFERENCES

[1] C. Anderson. The end of theory: The data deluge makes the scientific method obsolete. `http://www.wired.com/science/discoveries/magazine/16-07/pb_theory`, 2009.

[2] T. Arecchi. Phenomenology of consciousness: from apprehension to judgment. *Nonlinear Dynamics, Psychology and Life Sciences*, 15(3):359–375, 2011.

[3] T. Arecchi. Cognition and language: from apprehension to judgment. quantum conjectures. In G. Nicolis and V. Basios, editors, *Chaos, Information Processing and Paradoxical Games*, pages 319–339. World Scientific, Singapore, 2015.

[4] S. Barocas and A. Selbst. Big data's disparate impact. *SSRN eLibrary*, 2014.

[5] R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, 1950.

[6] L. Clark. No questions asked: big data firm maps solutions without human input. `http://www.wired.co.uk/news/archive/2013-01/16/ayasdi-big-data-launch`, 2013.

[7] K. Crawford. Think again: Big data. `http://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data`, 2013.

[8] N. Cristianini. On the current paradigm in artificial intelligence. *AICom*, in press.

[9] d. boyd and K. Crawford. Critical questions for big data. provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5):662–679, 2012.

[10] J. Haidt. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108:814–834, 2001.

[11] A. Halevy, P. Norvig, and F. Pereira. The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2):8–12, 2009.

[12] M. Hardt. How big data is unfair. understanding sources of unfairness in data driven decision making. https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de, 2014.

[13] R. Kitchin. Big data, new epistemologies and paradigm shifts. *BIG DATA AND SOCIETY*, 1(1):1–12, 2014.

[14] J. Lane, V. Stodden, S. Bender, and H. Nissenbaum, editors. *Privacy, Big Data, and the Public Good: Frameworks for Engagement,*. Cambridge University Press, 2014.

[15] P. Langley. *Elements of Machine Learning*. Morgan Kaufmann, San Francisco, CA, 1996.

[16] E. O. of President Obama. Big data: Seizing opportunities, preserving values. http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf, 2014.

[17] M. Pelillo and T. Scantamburlo. How mature is the field of machine learning? In M. Baldoni, C. Baroglio, and G. Boella, editors, *Proceedings of the Thirteenth International Conference on Advances in Artificial Intelligence*. Springer, 2013.

[18] M. Sebag. A tour of machine learning: an ai perspective. *AI Communications*, 27(1):11–23, 2014.

[19] I. Steadman. Big data and the death of the theorist. http://www.wired.co.uk/news/archive/2013-01/25/big-data-end-of-theory, 2013.

[20] L. Sweeney. Discrimination in online ad delivery. *Commun. ACM*, 56(5):44–54, 2013.

[21] S. T. T. Hey and K. Tolle. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond WA, 2009.

[22] L. Taylor, R. Schroeder, and E. Meyer. Emerging practices and perspectives on big data analysis in economics: Bigger and better or more of the same? *Big Data & Society*, 1(2), 2014.

[23] S. Watanabe. *Pattern Recognition: Human and Mechanical*. John Wiley & Sons, New York, 1985.

Friends, Robots, Citizens?

Stephen Rainey
Centre for Computing and Social Responsibility
De Montfort University
Gateway House, Leicester
+44 (0)116 207 7052
stephen.rainey@dmu.ac.uk

ABSTRACT
This paper asks whether and how an artefact, such as a robot, could be considered a citizen. In doing so, it approaches questions of political freedom and artefacts. Three key notions emerge in the discussion: discursivity, embodiment and recognition. Overall, discussion of robot citizenship raises technical, political and philosophical problems.

Whereas machine intelligence is hotly debated, machine citizenship is less so. However, much research and activity is underway that seeks to create robot companions, capable of meaningful and intimate relationships with humans. The EU flagship "Robot Companions for Citizens" project aims for "...an ecology of sentient machines that will help and assist humans in the broadest possible sense to support and sustain our welfare."[1]

This is a broad and ambitious aim, with a goal of making artefacts that can have genuine relationships with humans. This being so, in order to avoid merely creating highly interactive automata, the status of the robot must be carefully considered. Without significant public freedoms, for instance, the notion of a robot 'friend' would be a dubious one – as dubious as the notion of a 'willing slave', for instance. In a broad sense, these issues relate to the *politics* of robot kinship and sociality, perhaps specifically to civic epistemology. With a technological ideal of genuine human-artefactual kinship in the future, these political questions cannot be ignored. One approach to this problematic involves accounting for the robot citizen.

Categories and Subject Descriptors
K.4.0 [**General**]
K.4.1 [**Public Policy Issues**]: Ethics, Regulation
K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities
K.4.m [**Miscellaneous**]
K.5.0 [**General**]
K.5.2 [**Governmental Issues**]
K.5.m [**Miscellaneous**]

General Terms
Human Factors, Theory, Legal Aspects.

Keywords
Philosophy, Technology, Society

1. INTRODUCTION
The creation of a robot citizen cannot be achieved by purely technical means: citizenship throughout this paper concerns *taking an interest*. It is not clear that *any* amount of engineering can build this, nor have it be accepted as such by others. But there is an ineliminable technical part of the problem – not least in designing and building appropriate embodiment and cognitive faculties (it should be stressed that *cognition* is not purely facultative, though the faculties may be a necessary condition). However to truly achieve kinship with robots (as a shorthand for a variety of possibilities, e.g. machines, software, programmes), recognition must occur, or else every alleged companionship interaction would be dubious in the extreme. As with many aspects of human-human interaction (e.g. gender, race, occupation) the kind of human-robot recognition here required has a philosophico-political content that cannot be avoided.

This paper here hopes to develop a sketch of what would need to be the case for a robot to be considered a citizen, but not a manifesto and certainly not a guarantee.

The question of whether a robot *could* be a citizen is considered in terms of the conditions that would have to apply in order for a robot to share in place-making, where 'place-making' is an

[1] In the Strategic Partnership for Robotics in Europe Multi-Annual Roadmap (http://sparc-robotics.eu/about/) specific mention is made of "The ethical and social implications of social robots". In a broad conception of 'social', companionship and kinship between human and machine, human and programme, as well as inter-artefactual mutual reliances, partnerships, vulnerabilities and so on must be considered. Where genuine relationships are aimed at, discussions must go well beyond straightforward issues of human protection (http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5751970).

elaboration upon *merely* sharing space.[2] Citizens are explained as sharers of place, whereas *anything* can share a space with *anything* else.

The investigation will begin by looking at citizenship in very general terms, drawing upon Aristotle and Kant to substantiate the idea of 'taking an interest.' Drawing upon further philosophical thought from Searle and Habermas, the things in which citizens take an interest will be looked at. Finally, through the concept of embodiment, an exploration of *how* a robot could be thought of as taking an interest will be discussed.

2. CITIZENSHIP

The ability to contribute to the governance of one's political community is the notion central to citizenship in Aristotle's Politics Book III [2].

Aristotle makes a distinction between strict citizenship and qualified citizenship [2, pp1176ff]. The former can only be enjoyed by those free from service. This is to ensure that at any point citizens might be free to take part in governance. This is very much an active citizenship definition, one wherein the disposition toward political action is the marker of civic identity. Not all might enjoy the freedom to participate in governance that strict citizenship requires, notably in Aristotle's time slaves, women and foreigners. Foreigners could at least enjoy qualified citizenship. The point is none of these groups is thought of as being capable of contributing to governance, and so none can be politically active to the extent stipulated necessary for full civic identity

At the core, we can interpret beyond various ancient Athenian distinctions and say that citizenship is divided into at least two general groups which are citizens in a strict sense, and citizens in a qualified sense. For the citizen in a strict sense, the ability to take part in governance is a requirement. This in turn requires that those to be considered citizens must be free from impediments such as trades, poverty and service.

At least in principle, it would seem robots could easily fit the bill concerning freedom from trades, poverty and service. Were a robot to be constructed such that it had at least the semblance of free will, it would have no particular need to do any particular thing. That would rule out the need for a trade or service. Similarly, imagining a robot that was self-sufficient to the extent that many objects are, poverty would be no hindrance. It would not necessarily even be relevant. Yet, on this preliminary thumbnail sketch, this would not lead intuitively to an urge to partake in governance – where would be the impetus? This is one facet of the problematic which will be explored later, especially from section 4.

Aristotle's reasoning for granting unqualified citizenship to a particular group is that political society ought to exist for 'noble actions' and that these can issue only from a community, rather than from a mere alliance of various sorts of people. Aristotle's is a republican conception of citizenship, wherein participation or political agency is key. It assumes a fairly close agreement about ideas of the good life and about the various privileges of those involved in the community.

[2] The focus here isn't on robot rights, a short history of which can be found at http://www.roboethics.org/icra2005/veruggio.pdf (bullet-points at the end point to the sources of concern)

The republican model, in the shadow of Aristotle's Athenian ideal (Maybe typified by Florentine 'civic humanism'), may well be thought of as an impossible dream for modern, large, internally diverse and plural nation-states. If so, perhaps such republicanism can stand mainly as a critical standpoint from which to critique liberal political society. In fact, Kant can be read as hinting at something of a republico-liberal conception of the citizen, but on different grounds.

3. Kant

Kant suggests [9] that it is part of human nature that in society inevitable friction emerges as each individual seeks her own ends. This friction is offset by the claim that no single lifetime could feasibly accommodate the complete realisation of all of human beings' capacities. So, Kant supposes, the entire history of humanity is the arena wherein human beings' potential can be realised. This being so, politics is a necessary condition for human progress *per se* as it is politics that mediates the friction between the individual's plans and the progress of the community of all humanity. [5, p35]

In the context of an unfolding of humanity (of progress) and the necessity to act consistently with one's being an agent, one ought to do all one can to maximise the extent to which one can act and be unthwarted. From a historical point of view, social acting, on public reasons, is very important. Kant makes this point about law and freedom in terms of public and private reason, describing it as follows: Privately we must obey law, but always be ready publicly to challenge it:

> The public use of man's [sic] reason must always be free, and it alone can bring about enlightenment among men; the private use of reason may quite often be very narrowly restricted, however, without undue hindrance to the progress of enlightenment. But by the public use of one's own reason I mean that use which anyone may make of it as a man of learning addressing the entire reading public. What I term the private use of reason is that which a person may make of it in a particular civil post or office with which he is entrusted. [10]

At its most general, the importance of careful reasoning in terms of the public draws upon Kant's view on 'sensus communis'. This isn't 'common sense' as it would be known most widely, but rather is an *a priori* faculty of reasoning the denial of which would amount to a contradiction of agency in any given reasoner:

> ...under the sensus communis we must include the Idea of a communal sense, i.e. of a faculty of judgement, which in its reflection takes account (a priori) of the mode of representation of all other men [sic] in thought; in order as it were to compare its judgement with the collective Reason of humanity, and thus to escape the illusion arising from the private conditions that could be so easily taken for objective, which would injuriously affect the judgement. This is done by comparing our judgement with the possible rather than the actual judgements of others, and by putting ourselves in the place of any other man, by abstracting from the limitations which contingently attach to our own judgement. [11, §40]

The *sensus communis* is a form of individual judgement that takes into account others' partial ways of representing matters. The

point of this is to scrutinise *particular* judgements in this light of *general* human reason. This should avoid the individual, partial perspectives on matters that although personally compelling, could in general have a detrimental effect on any judgment made. From this constitutive principle of judgements in agents *per se* springs a motivation for taking an interest in public (or social) matters, taking care in that interest, and hoping others do likewise.

The republican elements of this, similar to Aristotle's view, are that shared community is important and an interest ought to be taken in it. It is to be cultivated privately by respectfully following the law that structures it, and publicly by challenging those tenets of the law inadequate to it. The liberal part of this conception has it that politics, *via* law, protects the citizen in terms of their personal freedoms. Citizenship sees its domain in the interaction between the private and public spheres. Citizenship here is to be thought of distributively, rather than as an aggregative notion.[3]

In short, drawing on Aristotle and Kantian thought, citizens must be a community, with a sense of community, and at least be disposed to taking an interest in the governance of that community. This will be referred to as 'place-making', over above mere sharing of space. The persistence of place-making comes through the fact that many citizens are interested in how things are run, how they could be better run, and the ends at which the running aims. From this starting point, place-making can now be elaborated upon, and its conditions laid out in order to determine what would need to be the case for a robot to take part in it, thereby grounding the chance of citizenship.

4. Place-making

The simple-looking question, "Where are you?" offers at least two potentially controversial interpretations. On the one hand, the question can be answered in terms of space. Answering in this way might involve the reporting of a set of co-ordinates relative to a grid. The question might also be answered in terms of place. This could involve the reporting of a more varied set of factors. These can be seen in an example from 'Mediterranean studies':

> If the classic work of Fernand Braudel (1949) tends to view the Mediterranean over the long term as a grand space or spatial crossroads in exchange, trade, diffusion and connectivity between a set of grand source areas to the south, north and east, the recent revisionist account of Peregrine Horden and Nicholas Purcell (2000) views the Mediterranean region as a congeries of micro-ecologies or places separated by distinctive agricultural and social practices in which connectivity and mobility within the region is more a response to the management of environmental and social risks than the simple outcome of extra-regional initiatives. [1]

[3] We can take from one of Kant's successors, Fichte, a parallel with his moral thinking, specifically the categorical imperative, and see the application for citizenship. The *I* and the *not-I* enjoy a mutual dependence, which is one reason for the necessity of treating others as ends on themselves – not to do so is to deny the mutuality of the *I* and *not-I* hence to deny oneself in a fundamental sense. Put briefly, to act against the other is not even to act but to be determined by a lack of understanding of what it is to be an actor, and agent, at all.

The first view is one that thinks of space in geometrical terms, whereas the second has a more holisitic view, drawing upon dynamic interests including the social. The latter is the place-making notion here put forward. An illustration might be helpful.

If Alice states that she is in Ireland, as opposed to in England, she means more than being simply further west than her London-based colleagues. Irish laws are different. Different customs operate. Different expectations mount when exploring Dublin as opposed to Dulwich, Dover (or Dresden, Darwin or anywhere else). This kind of difference between places over and above spaces is related to a holisitic notion of institutional reality, and social ontology, of the locations. This reality and ontology are the objects of interest for the engaged citizen.

4.1 Institutional Reality and Social Ontology

Institutional reality is a background to action for citizens. It offers a mode in which reasons can come which can warrant action in public. Reasons are required for action *qua* action and institutional reality provides a scheme from which are derivable desire-transcendent reasons (reasons not necessarily based in a metaphorical inner marketplace of competing personal desires) [14, p167ff] and a scheme *via* which desire can be articulated. Thus it presents a scheme from which action can result.

Searle describes how human beings have,

> ...the capacity to impose functions on objects and people where the objects and the people cannot perform the functions solely in virtue of their physical structure. The performance of the function requires that there be a collectively recognized status that the person or object has, and it is only in virtue of that status that the person or object can perform the function in question. [13, p7]

Status functions are concerned with the rules that constitute one thing as another in a context, e.g. a piece of paper as a payment in a shop, a person as a general in a war, a thrown towel as a submission. These things are *declared into existence* and so the normative power of speech acts makes something in the world that previously there was not. Searle here is flagging the *social* commitments and entitlements that constitute institutional reality. These commitments and entitlements serve to populate social ontology, moreover, as they are what create money rather than mere slips of paper; no-parking zones rather than mere tracts of land; a round of beers rather than a mere collection of liquids.

The collectivity of this is important. There is reciprocity at work without which these commitments and entitlements would become empty or disintegrate. There is then the question of who, or *what*, is taken as capable of ascribing, recognising and taking on social commitments of the sort that they can take part in the institutions of money, gambling etc. The question is actually familiar. Some people at some points are considered too young to take part in various institutions (in the UK, full time work for under 16s, 18 for gambling or drinking alcohol, 65 for receiving a bus pass). The roles are as much ascribed as the institutions: a bartender can sell beers she doesn't own, as she occupies a role given to her by the licensee, but she can't trade stocks in that role – there isn't a way in which my purchasing a Guinness can prompt Lloyds to remunerate the bartender or my portfolio to diversify. She isn't taken as a stock broker nor can she be both bartending for the bar and trading for Lloyds at the same time: illegitimacy in one or both roles creeps in. We can say from this that in occupying various roles and fulfilling them well, people mutually *enact* institutional and social reality.

The link between the enacting of institutional and social reality and place-making is worth highlighting at this point. It seems a clear advance beyond 'mere' space-sharing to be a constituent part of the possibilities for choice and action for oneself and others. In terms of the communitarian aspects of Aristotle and Kant mentioned above, this can also be seen in terms of taking an interest in the life of the community in that occupying these roles helps to constitute what can be expected in that very community.

4.1.1 The robot enacting social reality?

Whereas above it was asked from where the impetus to get involved in governance might arise for a robot, here the question becomes more complex:

- Could the robot be taken for something capable of ascribing status functions on one thing or another?
- To what extent could the robot *enact* institutional and social reality?

At the risk of littering the argument with too many rhetorical questions, this latest pair will remain hanging until the notion of *enacting* institutional and social reality is explored a little further. This can be done *via* Searle's notion of 'the Background' and the idea of 'civic nous'.

4.2 'The Background'

The role of social commitment as a structure to public action is ineliminable. There is a basic, possibly tacit, civic nous that guides interaction that can vary between place and place. In Searle, this is called 'the Background' [13, p135]. The Background is a set of mental states, not necessarily present to the mind at a given time, that sets up what the meaning of intentional states can be: they provide *the expected* from which divergence is noticeable (e.g. picking up an apparently heavy suitcase only to find it is a helium balloon shaped like a case. The surprise is analogous for the presence of the Background, though at any particular stage it wasn't called upon.)

Another way of thinking about it is as the set of justifications one would offer were one's routines to be interrogated (e.g. Why did you mime writing? I caught the waiter's eye – the bill has to be paid).

In terms of civic nous, the Background includes the capacity to recognise that standing on the left of a London Underground escalator constitutes a *faux pas*. The Background will underwrite spotting the error of an Englishman in Belfast inviting someone for a drink and actually having just one drink (rather than at least two). Whilst these sorts of examples might seem to indicate that the Background is simply a set of propositions, norms to be borne in mind, Searle argues [12, p156-7] that it is in fact not based in a mind-independent reality, but rather helps to structure the very reality that is inhabited. By way of another analogy it might be said that the flow of the Thames and that of the Soar look fundamentally alike. Invisible to the unaided eye is the bed that shapes and helps determine the unique way each river flows. In this analogy, 'the Background' is the riverbed, invisibly structuring the surface flow. It is something like a transcendental condition for the surface phenomenon – that *x* without which *y* could not be *y* at all – or of the Heideggerian 'immer schon' category that which is *always already* present.[4]

The London commuter doesn't hold in their mind any rules about escalators in order to avoid icy stares, having absorbed the fact that standing should only be on the right. The Belfast socialite doesn't consciously repeat the mantra that 'a drink' includes buying one back. Rather, each "...evolve a set of dispositions that are sensitive to the rule structure..." [13, p.145], where 'the rule structure' is the sets of social commitments collectively undertaken (without ceremony) in the context of the institutional reality in question.

Civic nous of this sort is the capacity to recognise contingency when the expected reality is deformed. Or again we might say that the Background manifests in social situations as that which gives content to surprise. This would suggest that contextualised cognition, this civic nous, is *embedded* in the institutional and social environment in which it appears, to allude once more to Heidegger, it is *always already* part of the logic of public being and public action, on which some more needs to be said.

4.3 Action

If we imagine a purely physical space of action, we can think of the laws of physics as the conditions for actions. The limitations of the body in contact with other surfaces are the limits of possibility here. The actions of the hypothetical dweller within the merely physical space are simply the instrumental interventions on the transcendent space. A physical space, if it is to be appreciated *as a place* in the sense here being used, will be a sphere of reasons besides.

Civic nous, the Background, collective status functions all come into play in a place. The intentions of the place-dweller, moreover, are structured according to the logic of the place: for instance, in wanting to buy something in London, Sterling is sought, rather than Euros. Places are shared spaces of action and so they come with a kind of a logical structure. Reasons can come in the sense of logical entailment or pragmatic presupposition, or more generalised warranty considerations regarding the sincerity and legitimacy of actions, among other things. I can command you to do something only if I'm warranted by being suitably superior in some regard. You can trust in my promise only if I am judged to sincerely undertake my obligation. Such warranty considerations occur within contexts like that of truth-preservation, wherein logical relations are of central importance, but operate on a less parsimonious conception of reasons than logical premises and inference rules. These reasons concern truth, truthfulness and normative rightness.

Assertions are obviously linked with truth, expressives with truthfulness (sincerity), and commands (etc.) with normative rightness (legitimacy or accountability). These three ways in which queries can be raised in conversation are themselves raised in Habermas' discussions of 'the validity basis of speech' as characterised in the late 1970s [13, p119].

The validity basis of speech is based upon the fundamental thought that in the very act of uttering, a speaker is claiming to be:

- giving [the hearer] something to understand

[4] See, for instance, Heidegger, M. (1996) *Being and Time*, trans Joan Stambaugh, Albany: State University of New York Press § 32, pp. 140-41

- making himself thereby understandable; and
- coming to an understanding with another person.

These are three 'world relations' taken to be implicit in speech action; fully successful speech acts must satisfy conditions of truth, sincerity and rightness (i.e. legitimacy according to some specifiable lights). [7, p75]

In the context of the robot citizen, the important question here is whether these world relations can be enjoyed by an artefactual agent. Or again, if such relations can be enjoyed, can they be recognised? The importance of these questions will be seen to hinge on some further details of the Habermasian account, and its relation to action in public, therefore to place-making over and above the mere sharing of space.

4.3.1 Further adventures in Habermas

Habermas supposes that validity claims in each of these three spheres can be raised and redeemed in communicative encounters, which amounts to raising and redeeming claims by means of argument. This being the case, validity in spheres beyond that of truth can be thought of as involving a notion of correctness appropriate to their own standards as truth is appropriate to claims of factual accuracy.

In this context, the phrase "validity claim," as a translation of the German term Geltungsanspruch, does not have the narrow logical sense (truth-preserving argument forms), but rather connotes a richer social idea—that a claim (statement) merits the addressee's acceptance because it is justified or true in some sense, which can vary according to the sphere of validity and dialogical context. [3]

By validity claims, then, is meant symbolic or explicitly made defensible propositions, sensitive to context; "A validity claim is equivalent to the assertion that the conditions for the validity of an utterance are fulfilled." [7, p38]

This is not necessarily the validity in which logicians are primarily interested, but rather must be taken to include in its scope the nuanced sense utilised above. Given we are in communication and not in some way merely noting one another's utterances, or in a therapy session or some other special type of interaction, we have to expect to be able to assume boundaries that themselves engender questions clustering around these three themes. That is to say, there are features of conversational interaction *per se* that we ought to be able to rely upon as underwriting expectations that can be shared by speaker and audience alike regarding the reasons that ought to be pertinent to their utterances. Reasons come in different flavours but can in general be requested owing to queries based in truth, sincerity and accountability.

In communication, then, what is of most interest is the role of rational compulsion as opposed to any other kind of motivation, such as fear of sanctions, for instance. The rationality associated with the simple securing of aims efficiently, instrumental (means: end) rationality, Habermas calls 'cognitive-instrumental' rationality. The concepts that would redeem validity claims in this context are simply those that, presupposing some goal, would with little fuss secure that goal. Habermas says that this much is true but goes on to stress the role of the criticisability of the knowledge claims in this area that is important but often overlooked.

The instrumental account presupposes knowledge of goals, circumstances and available means toward ends. Since with respect to each of these areas of presumed knowledge we can be mistaken, other people are apt to be able to show us that we are mistaken, perhaps by pointing out something that we've overlooked about the situation, for example.

Immediately, with this recognition, communicative rationality has expanded beyond the boundaries of mere means: end rationality. Now included in the list of presumed knowledge is knowledge of goals, circumstances, means toward ends and reasons and consequences (or ramifications). In short, the recognition of the criticisability of some presumed position opens a horizon for fallible propositional knowledge.

In acts of assertion, Habermas believes, the same knowledge is put to work as in teleological reasoning, but in a significantly different way. In action aimed at some goal, the actor can assess the rationality of their action alone and in silence. A criticisable assertion on the other hand must be rationally appraised in communication. It must be backed up with reasons or shown to be baseless with reference to another speaker's assertions in a public space of reasons.

A further extension of this simple realisation allows more candidates for rational appraisal than actions and assertions. If propositional contents can be rationally appraised on the basis of the redemption of the validity claims they raise, then other classes of expression too will be capable of rational appraisal based in the validity of the claims that they raise.

> In contexts of communicative action, we will call someone rational not only if he is able to put forward an assertion and, when criticized, to provide grounds for it by pointing to appropriate evidence, but also if he is following an established norm and is able, when criticized, to justify his action by explicating the given situation in the light of legitimate expectations. We even call someone rational if he makes known a desire or intention, expresses a feeling or a mood, shares a secret, confesses a deed etc. and is then able to reassure critics in regard to the revealed experience by drawing practical consequences from it and behaving consistently thereafter. [7, p15]

Thus, for an account of rationality larger than the mere means: end variety, we have to consider intersubjective communication in all its familiar forms. An intersubjective account of rationality has to include the possibility of validity of spheres such as those of sincerity, truth, efficacy, appropriateness, legitimacy etc. since these constitute real parts of communication. The spectrum along which validity claims can be raised and redeemed is thus much wider than merely goal-directed action and assertion.

4.3.2 Citizen-rationality and a public space of reasons

For any putative public agent, and so any citizen, this world of reason-giving and critique is a *sine qua non*. It is so owing to the requirement that place-making involves the holistic features noted from Mediterranean studies and the notion of public reasons. This is relevant to place-making as place-making requires taking an interest in the environment as a rationally structured space of reasons, as outlined in Searle's position. The critical potential contained within the dialogical account of citizen-rationality being outlined here makes the notion of acceptability important. Acceptability, in short, must be a reasoned acceptance, not an external determination, by a citizen of a norm, value, rule or what have you.

With this idea fleshed out by drawing upon Searle and Habermas, on the rationale given by Aristotelian and Kantian thought, we now have an account of what citizens ought to be thought of as taking an interest in (institutional reality and social ontology). We also have a way of understanding how they might take such an interest (Habermas' validity-theoretic account). What remains to be explored are the conditions that would have to obtain in order to grant access to this citizen-rationality and social institutional reality. In exploring this, some of the rhetorical questions raise so far can begin to be addressed.

5. Robot access to validity spheres?

The know-how brought to bear in being able to navigate the various contexts in which citizens routinely operate is *embedded* within the world in which they arise. Social action and social performances depend on facts about the world around us, including other citizens who are themselves *enacting* the institutional and social aspects of that world. That these terms arise in the manner that they do is interesting. Given so much of place-making (status functions, institutional reality, the Background, civic nous) is concerned with the contingencies of getting on in a shared, reason-providing environment which is enacted by those who inhabit it, it is highly probable that *embodiment* is central here too. This will be explored by way of the so-called '4Es' programme. This will begin to answer the questions raised above concerning the possibility, the impetus, that a robot could have for taking an interest in public life. This will be a beginning to understanding the conditions that would need to obtain for the robot to be understood as a place-making citizen.

5.1 Embodiment

Drawing upon the '4Es' research paradigm, it is possible to gloss a few relatively recent developments in thinking about cognition. These developments suggest that cognition is:

- Extended
    - the material vehicles underpinning cognitive states and processes can extend beyond the boundaries of the cognizing organism.
- Enactive
    - It depends on aspects of the activity of the cognizing organism
- Embodied
    - cognitive properties and performances can crucially depend on facts about our embodiment
- Embedded
    - cognitive properties and performances can crucially depend on facts about our relationship to the surrounding environment[5]

While these are intended to be read as insights to cognition, they can be deployed here in the context of this discussion of the robot citizen. The following sections will make the necessary connections to demonstrate this.

Thinking about the mere space-dweller, we can easily see parallels with various artefacts. For example, we might think of a robot mapping its environment by means of measuring paths of free travel and plotting obstacles so as to come to a geometry or a topography of the immediate area.[6] Were we to anthropomorphise here we could suppose the robot to be interested only in empirical truths concerning the environment. In considering the possibility of an artefactual citizen, however, it has to be asked whether and how a robot, programme or machine could get on with place-making.

This is now the opportunity to begin addressing the rhetorical questions raised earlier, *viz*:

- From where might the impetus to get involved in governance arise for a robot?

And

- Could the robot be taken for something capable of ascribing status functions on one thing or another? To what extent could the robot *enact* institutional and social reality?

What would need to be the case for a robot to meet the criteria for being a place-maker? Much of place-making is concerned with the contingencies of taking an interest in a shared environment, it seems likely that *embodiment* is central here.[7] Were citizens to be each of radically differing physical forms, the emergence of an institutional reality would not be clearly of interest to any particular individual. Nor might such an emergence be possible —

---

[5] Adapted from Ward, D., Stapleton, M., https://www.academia.edu/648508/Es_are_Good_Cognition_as_Enacted_Embodied_Embedded_Affective_and_Extended (November 2011):

"...the material vehicles underpinning cognitive states and processes can extend beyond the boundaries of the cognizing organism (Clark & Chalmers, 1998; Hurley, 1998; Clark, 2008).Cognition is enactive – that is, dependent on aspects of the activity of the cognizing organism (Varela, Thompson & Rosch, 1991; Hurley, 1998; Noë, 2004; Thompson 2007). Cognition is embodied – our cognitive properties and performances can crucially depend on facts about our embodiment (Haugeland, 1998; Clark, 1997; Gallagher, 2000). Cognition is embedded – our cognitive properties and performances can crucially depend on facts about our relationship to the surrounding environment (Haugeland, 1998; Clark, 1997; Hurley, 1998,). Finally, cognition is affective (Colombetti, 2007; Ratcliffe, 2009) – that is, intimately dependent upon the value of the object of cognition to the cognizer."

[6] For a brief overview see Thrun, S., *Robotic Mapping: A Survey*, http://robots.stanford.edu/papers/thrun.mapping-tr.pdf, 2002

[7] In terms of robot rights embodiment arises too. See for instance, a discussion on 'building in' ethics to robots mentioning humanoid forms and interactivity at:

http://link.springer.com/chapter/10.1007/978-4-431-54159-2_14. Also, in

http://www.i-r-i-e.net/inhalt/006/006_Veruggio_Operto.pdf

p.3 especially, the humanoid form is mentioned.

what might constitute general social norms for groups so diverse as to have radically different vulnerabilities and strengths?

Were citizens to be each of *radically* differing physical forms, the emergence of an institutional reality would not be clearly of interest to any particular individual. Where height, say, ranged randomly from millimetres to hundreds of meters, little sense could be made for, say, urban planning. Could a robot embodied as a dense, cubic kilometre of titanium, regardless of its faculties or apparent consciousness, possibly be understood as having interests in the environment comparable to putative fellow citizens? Or again, where a robot was embodied as a vast network of informational nods, ranging across galaxies, with an emergent consciousness, what sense could anyone make of it as a fellow burgher? It seems unlikely that such cases would permit the sort of *sensus communis* reasoning from Kant, or a comprehension of what validity claims could arise for such a being.

Another way in which embodiment reveals itself to be important in this context is in terms of the linguistic foundation to institutional reality that Searle points to and that Habermas elaborates. We can think of money as a promise, for example. Sterling notes actually state explicitly that they are promises from the bank to 'pay the bearer on demand.' Status functions in general are declared into existence and remain in existence through collective acknowledgment. The Background too can be seen as importantly linguistic, as the set of possible or counterfactual, justifications one would have given for an otherwise wordlessly performed act.

The particular way in which social beings are embodied plays a role in how and why they assign status functions the way they do, and so the institutional reality in which they act. The Background informs their mutual interactions like the terrain informs the way someone walks around. Civic nous has the impact on social action it does because it matters that another's social actions ought to be able to be anticipated and so personal actions not be perpetually frustrated.

Similarly with the case of the Background and civic nous, nowhere in particular is there a locus of this knowledge. There is a generalised pervasion of nudges, sways, insights and hints that constitute civic knowledge, that is, the knowledge of how to traverse institutional reality. From politeness on escalators, queuing for buses, paying bills, ordering beers in bars... laws, customs, habits, practices are nowhere codified once-and-for-all but rather they are more or less in any scenario to the extent that any given action is open to criticism or praise on how it matches up to this non-linear set of things.

Any artefact would seem therefore to need to be embodied in a comparable way to its social counterparts if it was to be considerable as a citizen. Any robot citizen would very likely need to be on a generally humanoid scale, with vulnerabilities similar to those of other humans.[8]

If civic, social or institutional reality is *enacted* by those whose relevant cognitive ability is *embodied* and reliant on being *embedded* amid details of the environment, then it is *extended*. The fact that this reality is extended makes it clear that it is public and up for grabs in a public way. No amount of navel-gazing reflection can arrive at a definition of what counts as this reality or its proper participants. Using the concept of recognition, we turn now to this last point.

5.2 Recognition

Could a public really detach itself from the view of the robot as servant? Could any given social group genuinely come to perceive the actions of robots as free in a robust sense? Given what has been said about embodiment just now, it could be guessed that a humanoid robot would stand a better chance than something thoroughly unlike a human in appearance. But it might also be guessed that even the most human-like robot would see diminished esteem upon the revelation of its artefactual nature. These are empirical questions, and themselves internally complex (i.e. is the possibility in question logical, practical, psychological etc.)

If the answers came in the negative, regardless of the actual capacities of the robot, none could ever be anything but a subject of oppression. Where recognition of agency is missing, there could be no chance of a full exercise of that very agency. In the republican senses of citizenship above this is the case owing to the unrecognised being unfree to take part in civic life, governance or the life of the community. In terms of Searle, the problem would be the robot not being taken as capable of enacting social reality. The suspicion of human citizens might be that the robot isn't experiencing 'the background' as the riverbed to their stream of action. Rather, something inauthentic might be suspected – behaviour in accordance with social norms read as rules. At best, the robot in these circumstances could enjoy only qualified citizenship, at worst be deemed imposter.

Bryson provides a perspective on robot identity that presses this negative line. [4] In this view, the robot is always, no matter how it is realised, an artefact directed by, and for the use of, human beings. The argument for this includes the claim that since human beings design, manufacture, own and operate robots, these robots are entirely the responsibility of human beings. This places them at the disposal of human beings, with at most the status of servant. Under no circumstances ought personhood or anything like it be attributed to the robot, on Bryson's analysis. To make such attributions would be to distribute incorrectly responsibilities and resources.

Certainly, in the area of interpersonal relationships this would be deeply problematic. Where a companion is sought, in the sense of a friend or partner, the freedom of the other is a necessary condition. Where that freedom is diminished in some way, relationships are possible but from a narrower base of, say, functional interdependencies. In the absence of robot freedom, robot companionship beyond such an interdependence is a non-starter.[9]

5.2.1 Servant machines

Bryson (*ibid*) offers a position paper and, perhaps as a result, the argument is somewhat unclear. A fourfold condition is deployed to underwrite the properly servile nature of the robot. The design, manufacture, owning and operation of robots raise different issues, especially with respect to responsibility. For example, where a robot's behaviour leads to, say, personal injury it is an

[8] One could imagine the argument running for other types of being in a similar way, such that humanoid scale mammals or artefacts would be problematic for them. Ditto softbots. The provision of a typology isn't the focus here, but could be a very interesting undertaking.

[9] And so the EU programme already mentioned would be a misguided novelty cf. http://www.robotcompanions.eu/

open question as to whether the responsibility for this lies in the design, manufacture, ownership or operation of the robot. Whether the designer, manufacturer, owner or operator is to take the blame for the bad outcomes is a serious question with potentially very high stakes.[10]

If the position stated in Bryson doesn't exhibit a genetic fallacy, discounting robot freedom on the basis of robot origins alone, its soundness might still be questioned. The part of the argument presented here[11] that robots cannot be more than servants states that:

1.) nothing designed, manufactured, owned and operated by human beings can be anything but for our use

2.) robots are designed, manufactured, owned and operated by human beings

3.) robots cannot be anything but for our use

Whilst this is a valid argument as it stands, assumption 1 seems to be controversial. A tremendous literature and research culture exists precisely to investigate the issues that would verify or falsify the proposition. It seems too quick to rely on this as assumption when much of what is at issue is contained within the very proposition. In fact, assumption 1 seems like a *refusal to recognise* robots as having a status beyond servant.

In fact, it seems evident that no matter the success or failure of the research programme aimed at clarifying the notions of assumption 1, it is not a guarantee that human beings would accept or reject robots as more than servants. The recognition of robots as citizens, or of any *x* as *y*, would in part involve what non-robots are willing to recognise as social or political involvement.

As has been argued elsewhere (in a different context), this cuts both ways. In the same way that machines could possibly be recognised as members of a community, "...so too might an unquestionably facultative being, of silicon, carbon, or anything else, be *excluded* or unrecognised where no such well of esteem exists." [6]

The refusal to recognise as valid institutional or social action subverts the status of the putative actor *regardless of their innate nature*. Action in context, recognised as such, is central to ascribing citizenship. From Searle's account, this active, context-sensitive dynamism is clear. Building upon Searle's account and drawing upon arguments above and the 4Es programme, it is possible to make a suggestion as to what would need to be the case for a robot to be recognised as a citizen:

Where the robot is *embodied* such that it has interests in the nature of public space, it can be considered as capable of taking part in social cognition *embedded* in details of the environment. In this context, it could be possible to recognise the robot as *enacting* various institutional or social roles that could constitute or enrich this *embedded* social cognition. The interplay of these extrinsic factors, open to recognition or not, would demonstrate the *extended* nature of institutional and social reality.

6. Conclusion

This paper pursued the following strategy in exploring what would need ot be the case for the possibility of a robot citizen:

It discussed citizenship in general terms, drawing upon a notion of 'taking an interest' and substantiated this with reference to Aristotle. An absence of dependence upon power is used in Aristotle as a *sine qua non* for strict, unqualified citizenship. Kant provided an even more general means of understanding the need for other-directed reflection where agency is at stake. Drawing upon Kant's account to make a political agenda, there is the sense that reason ought to constrain power, as private and public reason are contrasted. Between these two thinkers, a view of the individual and community is advanced, with a central place for freedom and reason.

The argument then discussed *in what* an interest should be taken, by the nascent citizen (once more in abstract terms). This was the 'shape' of institutional or social reality and this contextualised in a civic setting the sort of free and other-directed reasoning seen in the first step. Searle and Habermas provided material here which provided the objects for civic reasoning, but access to these object, or to this context, for the robot remained unresolved. How the robot citizen could gain this access was discussed in terms of embodiment, and the associated notion of recognition.

For the robot to be considered a citizen there is an onus on non-robots to recognise a robot citizen – robots can't be thought of as mere objects subject to arbitrary power. This is no small undertaking, especially when it is considered that many human beings still refuse such recognition for other human beings. An essential part of gaining recognition is the embodiment of the robot citizen

It was shown that embodiment was not just a simple device to garner esteem through mutual likeness between robot and non-robot. Rather, embodiment opens doors to enactivism, embedded social cognition and it acknowledges the extended nature of institutional and social reality. It provides a way in which to understand how things can come to matter to the robot citizen as they might matter to the non-robot citizen. It is a way in which the robot can be thought of as *taking an interest*. This lays the groundwork for the possibility of place-making beyond mere space-sharing, hence of citizenship.

7. ACKNOWLEDGMENTS

Many thanks to the ETHICOMP Panel reviewers, and the membership of the Centre for Computing and Social Responsibility at De Montfort University. Their helpful comments transformed earlier drafts of this paper.

8. REFERENCES

[1] Agnew, J, 2011 'Space and Place' in Agnew, J, and Livingstone D, (eds.) *Handbook of Geographical Knowledge*. London: Sage, 2011, extract at

http://www.sscnet.ucla.edu/geog/downloads/856/416.pdf

[10] See, for instance, the case of military robots: Taddeo, M., 'Information Warfare: A Philosophical Perspective' In *Philosophy & Technology*, March 2012, Volume 25, Issue 1, pp 105-120, 28 Jul 2011

[11] This isn't the only argument presented in Bryson's paper. An extended mind position is presented, for instance, urging a la Chalmers that robots can be thought of as extensions of our own minds. Perhaps so, but the assertion is too strong in being context-insensitive: friends, relatives, enemies and strangers could all be so thought of in the right context. It doesn't determine that robots can at most be servants.

[2] Aristotle, *Politics*, Book III, in McKeon, R., (Ed.) *The Basic Works of Aristotle*, Random House, NY, 1941

[3] Bohman, J, Rehg, W, 'Jurgen Habermas', in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/spr2008/entries/habermas/

[4] Bryson, J J, 2010 'Robots Should Be Slaves' in *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issue*s, Yorick Wilks (ed.), John Benjamins

[5] Dallmayr, F, 2001, *Achieving Our World: Toward a Global and Plural Democracy*, Rowman & Littlefield Publishers

[6] Erden, Y J, & Rainey, S, 2012 'Turing and the Real Girl', *New Bioethics: A Multidisciplinary Journal of Biotechnology and the Body* 18 (2):133-144

[7] Habermas, J, 1984 *The Theory of Communicative Action*, Vol. I, Polity Press

[8] Habermas, J, 1996 'What Is Universal Pragmatics?' in *The Habermas Reader*, (Outhwaite, W, Ed), Polity Press

[9] Kant,I, 1784, 'Idea for a Universal History from a Cosmopolitan Point of View', Translation by Lewis White Beck in Beck, L, W, 1963, *Immanuel Kant, On History*, The Bobbs-Merrill Co.

[10] Kant, I, 1784, An Answer to the Question: "What is Enlightenment?", https://web.cn.edu/kwheeler/documents/What_is_Enlightenment.pdf

[11] Kant, I, 1892 *Kritik of Judgement*, Translation by Bernard, J. H, London:Macmillan & Co

[12] Searle, J, 1983 *Intentionality: An Essay in the Philosophy of Mind* New York, Cambridge University Press

[13] Searle, J, 1995 *The Construction of Social Reality,* New York, Free Press

[14] Searle, J, 2001, *Rationality in Action*, Cambridge, MIT Press

Ethical, Legal and Social Concerns Relating to Exoskeletons

Dov Greenbaum
Zvi Meitar Institute for Legal Implications of Emerging Technologies
Interdisciplinary Center Herzliya
Yale University, Department of Molecular Biophysics and Biochemistry.
New Haven, CT
+1 917 365 1848
Dov.greenbaum@yale.edu

ABSTRACT

Exoskeletons, i.e., wearable robotics, are designed and built to amplify human strength and agility. In many cases, their purpose is to replace diminished or lost limb functionality, helping people regain some ambulatory freedom. As such, exoskeletons are particularly suited to help those with restricted mobility due to paralysis or weakened limbs. For all their promise, exoskeletons and other wearable robotics raise a number of ethical and social concerns that will need to be confronted by ethicists, the industry, and society as a whole. General social concerns relate to the personal and psychological impact on disabled individuals and their families. And as a society, we may need to reconsider able-ness, in light of these and other technological opportunities for overcoming our limitations. But that's only for those who can afford these machines: with exoskeletons costing as much as a luxury car, there are social justice concerns relating to access to this cost-prohibitive technology, as well as the eventual dependencies on such an expensive device. Ought insurers be required to purchase these for paralyzed individuals to significantly improve their quality of life; or are there competing interests and ideals that might support an insurer's refusal to invest in this technology? Some exoskeleton manufacturers, in conjunction with defense contractors, are reportedly pursuing military grade as well as industrial grade exoskeleton solutions. These solution enable soldiers and workers to perform longer and harder. In upgrading humans into quasi-machines, however, we run the risk of treating them more like machines than humans. In the workplace this may result in the overworking of an employee, in the military this could further dehumanize warfare and its very human actors. The prospect of augmenting otherwise healthy individuals (as distinct from treatment focused on achieving, sustaining or restoring health) raises further ethical concerns relating to human enhancement, an area fraught with slippery slopes. These issues are not only limited to our regular daily interactions, but also arise in sports, as the disabled (and now disgraced) Olympian, Oscar Pistorius, has shown us. There are no simple solutions for any of these issues, although many solutions may arise organically; for example, costs and access issues may be lessened as the technology becomes more widespread and cheaper. Other issues can be dealt with through well thought out regulatory solutions. But, for society at large, exoskeletons and other future human enhancements technologies raise much more longstanding and complex questions that will force us to redefine how we perceive humanity and self.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues, Social Issues, *ethics, privacy and regulation.*

K.5 **[Legal Aspects of Computing]** General and Government Issues.

I.2.9 **[Artificial Intelligence]** Robotics *Commercial robots and applications*

General Terms

Human Factors, Legal Aspects,

Keywords

Robotics, Exoskeletons, Ethics, Law, Social Issues, Disability Sports, autonomous

1. INTRODUCTION

Exoskeletons are an exciting emerging technology that promises, among other things, to provide re-gainable mobility to paraplegics. In this context, exoskeletons are, at their most basic, human-machine interfaces comprising robotics and computers, or more specifically, motors and sensors and software and novel algorithms that combine the former. While the concept of exoskeletons has been around for some time —see only the wide range of devices devised by our imaginations as represented in film ranging from Aliens (1986) to Avatar (2009)— the miniaturization of sensors, advancements in computing power and algorithms, innovation in battery technology and strong but light materials have all made what was once science fiction, a reality.

Given the potential of these technologies, exoskeletons are not only of interest to the disabled community where they provide the promise of walking, climbing and greater mobility, but they also present an exciting technology for the military, as well as for able bodied workers in industries requiring stamina, repetitive motion and hard labor.

This growing use of and interest in exoskeletons notwithstanding, there is a dearth of academic research on the ethical, social and legal aspects of this impressive technology. This is particularly

important in light of the general growing lag between the rate of technological innovation and the corresponding ethical, legal and social oversight of those technologies. Many of the ethical legal and social concerns raised herein will likely emerge sooner or later. As such, they ought to be dealt with, or at minimum, at least acknowledged and discussed before the technology becomes more integrated and enmeshed in society, and these concerns become more difficult or even impossible to overcome.

1.1 Goals of the Paper

This paper will be an attempt to reverse this research gap by moving the policy engagement upstream; instead of regulating as a reaction to technology, this paper aims to provide anticipatory expert opinion that can provide regulatory and legal support for this technology, and perhaps even course-correction if necessary, before the technology becomes ingrained in society.

As such, this paper intends to highlight many of these non-trivial issues. The paper will look to ethical, legal and social issues separately, although in some instances, many of these issues may overlap, and have repercussions in other areas. However, while this paper intends to provide a broad overview of the issues, the concerns raised represent only some of the pertinent issues and are not intended to be an exhaustive list.

1.2 Human Enhancement

What is and what is not human enhancement is a central concept to many of the legal, social and ethical issues associated with exoskeletons. Unfortunately, the definition of this term remains ambiguous and is non-trivial. Currently there is no consensus as to what ought to be considered human enhancement per se and what is not.

Thus whereas researchers can generally agree that human enhancement comprises the extraneous, non-natural providing of skills or abilities beyond those typical to the species, it's not clear, for example, at what point an added-on tool becomes more than just an extraneous tool and becomes an incorporated enhancement. More specifically, at what point does an exoskeleton become sufficiently integrated (either internally, or even externally both physically and/or through a brain interface[1]) such that it is an actual extension of the individual, and an enhancement thereof, and not simply just an external tool.

Some have argued that perhaps an "always on" feature of a tool, changes it from an extraneous tool to an integral and integrated component of the individual.[[2]] However, in the example of exoskeletons, battery life limitations make the fulfillment of this criteria unlikely. Moreover, under an "always on" theory of human enhancement corrective lenses might also be considered an enhancement.

An additional/alternative criteria in defining enhancement may relate to the distinction between therapeutic and non-therapeutic manipulations. Under this criteria, non-therapeutic changes or alterations could be considered enhancements, (or are more likely than therapeutic changes to be considered enhancements), whereas most if not all purely therapeutic changes would fall out of the ethically problematic realm of human enhancement. However, this criteria is also problematic as it is not clear whether therapeutic changes ought to be limited to regaining the individual's status quo (e.g., on par with the average of the species), or whether they should include therapeutic changes that exceed the status quo.

Following the corrective lenses analogy, LASIK or similarly eye altering surgery which raises the individual's eyesight back to the status quo would not be considered enhancement. But, under this definition, the commonly performed and widely accepted Tommy John Surgery, e.g., ulnar collateral ligament (UCL) reconstruction wherein ligaments from other parts of the body are used to reconstruct parts of the elbow, might be considered human enhancement ;[3] evidence indicates that baseball pitchers with declining skills prior to surgery see an improvement in their skills post surgery,[4] with some even calling it surgical doping.[5]

This last criteria in defining what is and what is not enhancement is further problematic as it is arguably biased against those who were born with a disability relative to those who's disability came later in life. For those born with a disability changes to body would arguably be an enhancement as they raise the individual above her status quo, whereas the same changes for an individual who became disabled may simply return that individual to the status quo.

While there are no easy answers in defining human enhancement, a definition is eventually necessary for a legal analysis; whether something like an exoskeleton falls under the rubric of human enhancement remains a paramount issue in devising regulation of exoskeletons.

In particular, scholars tend to fall into three camps in assessing the level and nature of necessary regulation. [6] Some argue that more than a base minimum would only serve to disincentivize future technological developments and would clash with the natural right to control ones' own body.[7] These transhumanists argue that human aspects of freedom and autonomy demand that we be able to enhance at will.[8] Others argue that the potential side effects and social upheavals that could result from human enhancement technology requires strong regulations, if not even a moratorium on research in this area until we can work through all the problems. [9] In between these two poles are those who argue for regulation developed in light of the inalienable rights to control our own bodies.[10] Andy Miah has described the sometimes bizarre and illogical position of these last two groups: "We embrace all those enhancements that we have deemed a reasonable extension of natural ability and we carefully regulate those that we haven't."[11]

1.3 Are Exoskeletons human enhancements

The lack of definition as to what is or what is not human enhancement notwithstanding, it would be hard to state that exoskeletons are per se human enhancement: Exoskeletons have varied purposes and integrate with the body differently depending on the manufacturer and/or goals, For example, an exoskeleton that allows a paraplegic to regain some motor skills, arguably ought to be viewed as very different than an exoskeleton that is used by a soldier to obtain extra fighting skills in the eyes of the law.

Whether or not Exoskeletons are considered enhancements, exoskeletons however, are likely, in all their forms, robots, i.e., a physical machine that obtains data from the environment, processes that information and then interacts with the environment based on that data: In Caro's formulation, one that "senses thinks and acts."[12]

And, like being defined as a human enhancement, being defined as a robot brings its own baggage of robotics exceptionalism, as it has introduced a systemic change to the law in dealing with this technology with dozens of US states having robot-specific laws.[12]

1.4 Exoskeletons Currently in Development

Currently there a number of companies working on developing exoskeletons for both the military and consumer use, both therapeutic and non-therapeutic. Key stakeholders in this area are Cyberdyne's Hal, Ekso Bionics, Argo Medical's ReWalk Robotics, Parker Hannifin's Indego and Rex Bionics.

ReWalk, the result of an Israeli company, was the first exoskeleton to obtain FDA approval for the use of their technology for paraplegics, is relatively expensive at around 70,000 dollars per device. This high cost notwithstanding, there are a number of institutions worldwide that provide access to these exoskeletons for therapeutic usage. A number of clinical trials are also underway to examine the usage of this technology.

Ekso Bionics, in conjunction with Lockhead Martin, developed a number of non-therapeutic exoskeletons including, HULC (Human Universal Load Carrier) with military usages, ExoHiker, which helps hikers carry large loads, Exoclimber, specifically designed for stairs and slopes and eLEGS, (Exoskeleton Lower Extremity Gait System) which is a hydraulically powered system that could allow paraplegics to stand and walk with additional support.

Cyberdyne, a Japanese company with an ominous name has a line of exoskeleton robots that provide both therapeutic and non-therapeutic usages. The therapeutic uses of these devices include uses for individuals with brain and mobility disabilities and the non-therapeutic uses including eldercare and worker assistance devices.

Rex Bionics, part of Edison Investment Research Limited. Rex is focused on rehabilitating patients with spinal injuries and disabilities relating to stroke or Multiple Sclerosis with a focus on both home use and rehabilitation institutions.

1.5 Exoskeletons in Popular Culture

The ethical legal and social concerns relating to exoskeletons are arguably exacerbated by the use of exoskeletons in popular culture, particularly in film where more often than not, they provide the user with extraordinary abilities. Lists of exoskeletons in film abound online and include films dating back to the 1950's. Most of these suits grant their wearers strength, agility and other powers. [13] most are associated with aggressiveness and warfare, few, like the Caterpillar Power Loader in the 1986 movie Aliens is designed for picking up heavy objects. This public perception of exoksleons as fighting machines potentially confounds the man other positive uses of such technologies.

2. ETHICAL CONCERNS

As described above, augmenting humans is rife with concern. And while it can be easily justified in some situations, e.g., for therapeutic purposes, in others their use is typically ethically more problematic.

However, in addition to just strapping on an exoskeleton for no particular reason, there are a number of defined opportunities for non-therapeutic, dual-use-like enhancement that might be particularly problematic; for example in in sports, heavy industry and military applications. Here the ethical dilemmas are even more pressing.

2.1: Dual Use

With a number of exoskeleton manufacturers focusing on the industrial and military uses of the technology, we run the risk of dehumanizing our workers and our soldiers that are strapped into exoskeletons. For example, in the case of industry, the use of exoskeletons in areas requiring heavy repetitive lifting, managers and others overseeing the workers may overlook the human components and needs of their workers, seeing them only for their enhanced mechanical abilities that the exoskeletons provide them. As will be discussed later this may also have legal implications.

2.1.1 Industrial Use

At minimum when exoskeletons are incorporated into industry, from construction to manufacturing, to even geriatric care providers, rules and regulations ought to be promulgated that protect the workers from being dehumanized and overworked.

2.1.2 Military Use

With the prospect that soldiers might be upgraded uparmored and otherwise enhanced by exoskeleton technology comes the risk that not only will the enemy fail to see the soldiers as humans, a particular problem for our soldiers and a propaganda coup for the other side, but so will the soldiers commanding officers.

One voiced concern is that commanding officers might expect their enhanced soldiers to be able to work harder and longer, with more consideration for their robotic side and perhaps with lesser concern for their mental wellbeing as a result of this harder work. Additionally, commanding officers, in seeing even a little less humanity in their soldiers, might be more likely to send their charges into dangerous or difficult situations, situations that they would have avoided had the soldiers not been mechanically enhanced. Soldiers in armies tend to also have fewer rights than civilians; regulations may be necessary to limit the ability of the military to test exoskeletons on soldiers without the use of informed consent and other legal safety nets.

Moreover, according to those theories that war is supposed to be as horrible as possible to disincentivize combat between parties, the mechanization of the soldier plays into that mindset, making war worse. Additionally, the enhancing of soldiers makes the political cost of war less, as its assumed that mechanized soldiers will be less likely to become politically costly casualties. In any event, its likely that the eventual use of exoskeletons in battle will necessitate a rewriting of some rules of engagement.

Finally, in general taxpaying citizens supporting scientific innovation may be concerned with the dual use nature of the technology wherein the funded research may have initially been intended to create life enhancing technologies and only later being coopted into military and non-therapeutic uses.

2.1.3 Use in Sports

In addition to the obvious problematic areas of dual use, there are additional ethical concerns raised with the eventual incorporation of this and related technologies into amateur and professional sports and the social disruption resulting from the incorporation of this technology. (Some have already argued for multiple leagues in sports including separate leagues for the enhanced and not-yet enhanced.[14])

Eventually, lines will need to be drawn to determine what amounts to illegal or unfair augmentation and what remains fair enhancement by exoskeleton, if any. Here fairness might take into account any harms that might be caused to the athlete as a result of the technology, the dehumanizing or superhumanizing of said athlete, the virtuousness of the enhancement, and whether or not the resulting enhancement is against the practically undefinable concept of the spirt of the sport. [15]

The issue of enhancement in sports is not necessarily a novel issue, as every time a new technology arises, the sports authorities

need to determine whether that technology will be allowable. For example, whereas hyperbaric chambers and tents remain allowable, some bathing suits that aim to mimic shark skin are not.

In some instances, precedents may have already been set, for example with regard to exoskeletons and their use in marathons. A number of marathons have already allowed disabled athletes to run using these technologies. While currently, the use of the technology doesn't threaten the standings of the top athletes, but will the marathons reconsider their adoption of these technologies when records are threatened? Or will we see separate categories of runners, in addition to gender, enhanced and non-enhanced.

In examining this question it is important to recognize that whereas conventional wisdom sees our top athletes as the product of blood sweat and toil, in reality, most if not all are born naturally genetically enhanced to compete, including for example, longer limb length for some top swimmers, or greater oxygen carrying capacity for certain bikers.

3. SOCIAL CONCERNS:

3.1 Access

In addition to these ethical concerns, there are a number of social concerns. For example, currently the technology for enhancement of the disabled is somewhat costly, limiting access to those few who can either afford to purchase access to the technology or those lucky enough to have health insurance plans that will pay for the costs associated with using this technology.

This goes to the much more difficult question regarding whether the disabled have a right to technology that returns them to an equal playing field with their peers. Does human dignity demand that we do all that we can for those less fortunate than ourselves? Can the disabled argue that they have a right, under their governments to access this technology at a reasonable and affordable price?

This discrimination of access, based solely on ability to pay without recognition for the type of injury or the health benefits raises non-trivial social justice concerns in addition to ethical concerns relating to the role that expensive exoskeletons play in actively further relatively disadvantaging those who are disabled but cannot afford this technology. While we are mindful that with regard to all areas of human enhancement, fair distribution of the technology is not necessary equitable distribution of technology. Moreover, non-equitable distribution of the technology, as described above with the dual use nature of the technology can create a market wherein eventually economies of scale will result in the technology being more affordable for everyone.

In discussing the social aspects of access, ought health care providers to pay for everyone to have access? How should insurers decide who does or does not get access to this technology.

Moreover, as a society, perhaps we be promoting more dual use of this technology, if for no other reason than that economies of scale might reduce the cost to use and/or obtain an exoskeleton for disabled individuals.

3.2 Dependency and Withdrawal

There may be concerns that the availability of exoskeletons will create a dependency on the technology, and a limited availability will lead to withdrawal like symptoms, wherein disabled individuals who may have relied on the technology, may exhibit psychosocial withdrawal-like symptoms when they lose access, either because of scarcity or because they can no longer afford access.

3.3 Defining Ableness and Disability

In addition to the social justice concerns regarding access and dual use of the technology, they are additional concerns relating to the definition and reassessment of the definitions of ableness and disabilities. With the prospect that humans can be augmented with integrated exoskeletons and other prosthetics, we may need to reassess what defines ableness and disability and in particular, whether an individual augmented with an exoskeleton such that they regain the ability to walk and/or otherwise be mobile to an some degree, an equal degree, or perhaps in the near future, to a greater degree than those without the exoskeleton, is still disabled.

A simple minded comparison might be a comparison between individuals with 20/20 vision, individuals with glasses, individuals who have undergone LASIK surgery to regain or even further enhance their vision and individuals who have incorporated contact lenses that provide telescopic vision and/or ight vision.[16] Are the individuals with glasses impaired in comparison to those with natural 20/20 vision? What about those who wear contact lenses? What about the individuals who have undergone lasik surgery to regain 20/20 eyesight, are they similar to or different than individuals with glasses in comparison to those without.

This comparison is not without precedent. Under the American's with Disabilities Act (ADA, 1990), perhaps the preeminent civil rights statute for the disabled in the United States, and described by James Brady in a New York Times Editorial as necessary statue for people with disabilities - the largest minority in the U.S. [who] were left out of the historic Civil Rights Act of 1964,[17] the need for corrective lenses is not per se, a disability under the ADA: "In enacting the Americans with Disabilities Act of 1990 (ADA), Congress intended that the Act "provide a clear and comprehensive national mandate for the elimination of discrimination against individuals with disabilities" and provide broad coverage;"[18]

As per the U.S. Equal Employment Opportunity Commission (EEOC) website, "Ordinary eyeglasses or contact lenses" – defined in the ADAAA and the final regulations as lenses that are "intended to fully correct visual acuity or to eliminate refractive error" – must be considered when determining whether someone has a disability. For example, a person who wears ordinary eyeglasses for a routine vision impairment is not, for that reason, a person with a disability under the ADA. The regulations do not establish a specific level of visual acuity for determining whether eyeglasses or contact lenses should be considered "ordinary." This determination should be made on a case-by-case basis in light of current and objective medical evidence."[19]

In Sutton v. United Airlines wherein the US Supreme Court determined that a definition of disability ought not be adjudicated "in their hypothetical uncorrected state—is an impermissible interpretation of the ADA. Looking at the Act as a whole, it is apparent that if a person is taking measures to correct for, or mitigate, a physical or mental impairment, the effects of those measures—both positive and negative— must be taken into account when judging whether that person is "substantially limited" in a major life activity and thus "disabled" under the Act"[20]

Later, under the 2008 amendments to the ADA (ADAAA), signed into law by George W. Bush, 18 years after his father George H. Bush signed the ADA into law, Congress made a conscious effort

to broaden the term disability, heretofore narrowed by Sutton and its progeny. [21] The ADAAA is designed to "reject the requirement enunciated by the Supreme Court in Sutton v. United Air Lines, Inc., 527 U.S. 471 (1999) and its companion cases that whether an impairment substantially limits a major life activity is to be determined with reference to the ameliorative effects of mitigating measures;"[22]

Under the above mentioned 2008 amendments to the Act, regulatory agency guidelines to the contrary[23] were codified such that: mitigating measures, i.e., those that "eliminate or reduce the symptoms or impact of an impairment [including] medication, medical equipment and devices, prosthetic limbs, low vision devices (e.g., devices that magnify a visual image), hearing aids, mobility devices, oxygen therapy equipment, use of assistive technology, reasonable accommodations, and learned behavioral or adaptive neurological modifications [,may not] be considered when determining whether someone has a disability … In other words, if a mitigating measure eliminates or reduces the symptoms or impact of an impairment, that fact cannot be used in determining if a person meets the definition of disability. Instead, the determination of disability must focus on whether the individual would be substantially limited in performing a major life activity without the mitigating measure. "[24]

Notably, however, " the positive or negative effects of mitigating measures [may] be considered when assessing whether someone is entitled to reasonable accommodation or poses a direct threat." [24] As such, employers "can take into account both the positive and negative effects of a mitigating measure. The negative effects of mitigating measures may include side effects or burdens that using a mitigating measure might impose" [24] As such, "if an individual with a disability uses a mitigating measure that results in no negative effects and eliminates the need for a reasonable accommodation, a covered entity will have no obligation to provide one." [24]

And while an employer cannot require "an individual to use a mitigating measure. However, failure to use a mitigating measure may affect whether an individual is qualified for a particular job or poses a direct threat." [24]

Considering these regulations in the context of an exoskeleton, while an employer cannot ignore the fact that a person is disabled simply because they employ an exoskeleton, and while an employer cannot force an employee to use an exoskeleton, the use of an exoskeleton by an employee may act as a mitigating measure sufficient to find that the employee is not in need of any reasonable accommodations by the employer. Further, as the cost of exoskeletons go down, one could conceive of a time in the near future wherein an employee could demand the use of an exoskeleton as a reasonable accommodation by the employer.

In light of the mixed response of the ADA to mitigating technologies, the use of an exoskeleton further confounds the self-identification of individuals as disabled or not disabled. Like hitech prosthetic limbs that nearly mimic true function of a lost limb, exoskeletons may soon unobtrusively mimic the true function of a limited-function limb leading some people to self-identify as disabled, and others to perhaps self-identify as not disabled. In all likelihood this will create substantial confusion in the general public and particularly in the service industries that given this scenario, might struggle to assess what level of service is necessary for these individuals.

4. LEGAL ISSUES

Social issues of ableness and disability reach into legal issues, as described above. In addition to issues relating to disability, there are a number of other issues relating to the law.

4.1 Exoskeletons in Court

As with all new technologies, in the US jury system, lawyers in the early cases will have the opportunity to establish the necessary metaphors to properly frame the technology to suit their case. In these early cases, unfavorable precedent could be set —bad facts make bad law— to shoehorn all exoskeletons into one metaphor or another. [25]

4.1 Exoskeletons in Criminal Law

Criminal law requires that the actors have bad motivations for their actions. With exoskeletons, the motivation analysis may be confounded by the autonomous or semiautonomous nature of the devices and the nature of the human-machine interface.

4.2 Exoskeletons in Tort Law

In tort law, courts look to, among other factors, the foreseeability of the tortious result of an action in assessing the negligence of the actor. With regard to exoskeletons, the interaction between human, motors, sensors and software may not always result in forseeable results. This is all the more complicated by autonomous and semi-autonomous exoskeletons that may interact with the environment irrespective of the intentions of the user. Further confounding these issues, concerns may arise when the machine-human interface includes direct neural connections between the user and the device, wherein unconscious or subconscious intentions may be translated into actions by the exoskeletons, those actions may result in a tort.

Additionally, the use of the common law theory of res ipsa loquitor wherein the courts acknowledge the imbalance of information between the tortfeeser and the victim, may become unmanageable in cases of exoskeletons wherein the multiple stakeholders associated with the exoskeleton, including the manufacturer, the programmer, the user, among others, makes it unlikely that anyone has a good handle on the information.[25] This is particularly the case under the Restatement (Second) of Torts, wherein § 328D outlines a process for finding negligence by the tortfeeser: determining whether the accident is one typically the result of a negligent action, and more problematic in the case of exokseletons, that the defendant had exclusive control over the instruments that were the proximate causes of the tort. In the case of exoskeletons, it may be difficult to infer that a user of an exoskeleton had exclusive control over the autonomous or semiautonomous robot. Notably, though The Restatement (Third) of Torts, § 17, leaves out the exclusive control element.

4.3 Exoskeletons in Product Liability Law

In general, in product liability law we often look to strict liability, finding the producer of a device liable irrespective of their negligence. Moreover, in some instances there are different criteria for liability depending on whether the faulty device is a medical device or a non-medical consumer device. At this point, FDA approval for the device notwithstanding, its not clear how tort law will treat faulty exoskeletons.

In some areas of product liability the law has imposed strict liability on faulty products. However, strict liability falls away in some areas of technology, including software, were society has come to acknowledge and expect glitches and software bugs.[26]

In the case of exoskeletons, it is unclear whether courts will enforce strict liability, as is common in other machine-software devices, such as cars, or whether a different standard will be set.

4.4 Exoskeletons and Privacy

Exoskeletons by design may collect data on the user. This data collection may be necessary for product feedback and/or medical necessity. For example, exoskeletons may collect location information, usage information, neural input information, vitals data and other private information relating to the user. Regulations would need to be developed, not only to standardize this data collection so that it can be useful cross platforms, but all to enforce encryption and/or other levels of protection when the data is at rest, data in use and data in motion.

4.5 Exoskeletons and Workers Compensation

Under standard Workers Compensation theories, employers pay workers compensation to injured employees in exchange for legal leniency if an employee becomes injured in their place of employment, potentially due to a fault of the employer. If and when workers begin to use exoskeletons in the workforce, workers compensation for employees injured while wearing an exoskeleton may be limited, but the employee may have recourse in going after the producer of the exoskeleton.

4.6 Exoskeletons and workers' rights

Currently many workers have sets of rights that limit their work hours and that sets wages, among other worker related rights. It is unclear how exoskeletons may change the amount of time the law is willing to let employers work their employees, and whether compensation may be different for employees that use exoskeletons and those that do not use exoskeletons.

5. CONCLUSIONS

Although the exoskeleton industry is in its infancy, it is obvious that there are a number of ethical, legal and social concerns that must be acknowledged and maybe even dealt with before the technology becomes entrenched and bad precedent creates legal, social and/or ethical realities that might hinder future development of the technology and/or harm the users of the technology.

6. REFERENCES

[1]Demetriades, A. K., Demetriades, C. K., Watts, C., & Ashkan, K.. Brain-machine interface: the challenge of neuroethics. *The surgeon*, 8(5), 267-269, October 2010.

[2]Allhoff, F., Lin, P., Moor, J., & Weckert, J. Ethics of human enhancement: 25 questions & answers. *Studies in Ethics, Law, and Technology*, 4(1), February 2010.

[3]Miah, A. Rethinking enhancement in sport. *Annals of the New York Academy of Sciences*, 1093(1), 301-320, December 2006.

[4]Gupta, A. K., Erickson, B. J., Harris, J. D., Bach, B. R., Abrams, G. D., San Juan, A., . & Romeo, A. A. Performance and Return-to-Sport after Tommy John Surgery in Major League Baseball Pitchers. *Orthopaedic Journal of Sports Medicine*, 2(1 suppl), 2325967114S00022, December 2014.

[5]Rodenberg, R. M., & Hampton, H. L. (2013). Surgical doping: a policy loophole?. *International Journal of Sport Policy and Politics,* 5(1), 145-149, March 2013.

[6]Allhoff, F., Lin, P., & Steinberg, J Ethics of human enhancement: an executive summary. *Science and engineering ethics*, 17(2), 201-212, June 2011.

[7] Harris, J.. Enhancing evolution: The ethical case for making ethical people. Princeton: Princeton University Press, 2007.

[8] Kurzweil R The singularity is near: when humans transcend biology. Viking Penguin, New York, 2005.

[9]Furger, F., & Fukuyama, F. Beyond bioethics: A proposal for modernizing the regulation of human biotechnologies. *innovations*, 2(4), 117-127, Fall 2007.

[10]Greely, H. T. Regulating human biological enhancements: Questionable justifications and international complications. *UTS L. Rev*., 7, 87 2005.

[11] Miah, A. Enhanced Athletes: It's Only Natural. *Washington post*, August 2008.

[12]Calo, R.. Robotics and the Lessons of Cyberlaw. California Law Review, 103, 2014-08, June 2015

[13] Sofge, E A History of Iron Men: Science Fiction's 5 Most Iconic Exoskeletons, Popular Mechanics, April 2010 available online at http://www.popularmechanics.com/culture/movies/a5523/scifi-most-iconic-exoskeletons/

[14]King, M. R. A League of Their Own? Evaluating Justifications for The Division of Sport into 'Enhanced'and 'Unenhanced'Leagues. *Sport, Ethics and Philosophy*, 6(1), 31-45 February 2012.

[15]Miah, A. Rethinking enhancement in sport. *Annals of the New York Academy of Sciences*, 1093(1), 301-320, December 2006

[16] Hanlon, M, Electronic Contact Lens promises bionic capabilities for everyone. *Gizmag.* 21 January 2008. Available online at http://www.gizmag.com/electronic-contact-lens-promises-bionic-capabilities-for-everyone/8689/.

[17] Brady JS. Save Money: Help the Disabled, Editorial, New York Times, August 29, 1989 available online at http://www.nytimes.com/1989/08/29/opinion/save-money-help-the-disabled.html

[18] ADA Amendments Act Of 2008 (ADAAA) PL 110-325 (S 3406) September 25, 2008 (ADAAA 2008) §2(a)

[19] http://www.eeoc.gov/laws/regulations/ada_qa_final_rule.cfm

[20]Sutton v. United Air Lines, Inc., 527 US 471 - Supreme Court 1999

[21]Sauer, E. ADA Amendments Act of 2008: The Mitigating Measures Issues, No Longer a Catch-22, *The. Ohio NUL Rev.,* 36, 215, 2010

[22] ADAAA §2(b)(2)

[23]29 CFR pt. 1630, App. § 1630.2(j); 28 CFR pt. 35, App. A § 35.104 (1998); 28 CFR pt. 36, App. B § 36.104.

[24] http://www.eeoc.gov/laws/regulations/ada_qa_final_rule.cfm

[25]Calo, R.. Robotics and the Lessons of Cyberlaw. California Law Review, 103, 2014-08, June 2015

[26]Calo, M. R. Open Robotics. Maryland Law Review, 70(3), 571, May2011

Japanese cultural and ethical *Ba* (locus) as the place of new sources for technological and social innovation as well as for ethical discussions on robots and life in the information era

Makoto Nakada
University of Tsukuba
1-1-1 Tennoudai
Tsukuba Ibaraki Japan
+81 29 853 4037
nakada@japan.tsukuba.ac.jp

ABSTRACT

In the first part of this paper, the author tries to analyze the crisis that Japanese people of today face, citing various statistical data provided by the Japanese government and the research groups interested in this problem including the author's research data themselves. This crisis means that Japan of today has to deal with a lot of serious situations such as the economic stagnation measured by the GDP growth, loss of export capacity of high-tech industries, the high suicide rate, the decline of the local economy and so on. The important thing regarding this Japan's decline is the fact that these serious situations have started to appear in that time of so-called 'informatization.'

In the second part of this paper, the author attempts to find out the factors related to these serious situations in Japan of today. In the author's view, one of the most important factors affecting Japan's culture, society and industries might be a loss of 'depth' of Japanese culture and society, or (loss of)awareness of the place (*Ba*) from which people can extract various meanings of life coming from Japanese cultural traditions, senses of 'oneness' or the experiences in the past history.

The phenomenon of loss of cultural depth or loss of awareness of *Ba*-related meanings is one of the most serious problems in Japan today and this leads to serious problems such as high suicide rate in Japan since 1998. But on the other hand, at least at a latent level, it was found through the author's own previous researches [11, 12, 13, 14, 15] that Japanese people still tend to show strong sympathy for the meanings related to *Seken*, i.e. the traditional aspect of Japanese culture and society or *Ba* with traditional cultural, ethical, existential meanings.

And finally, the author wants to focus on 'another depth' which seems to be related to traditional ways of understanding of oneness(*Ichinyo*) in Japan, i.e. the world view putting emphasis on the situations of undifferentiation of things, persons, events, nature.

In addition to this kind of traditional oneness, we have to carefully see some sort of 'artificial oneness' including emergence of autonomous robots or phenomena related to 'rubber hand illusion,' 'mirror box therapy' and so on: 'artificial oneness' happening in the artificial environments in the information era.

In this sense, the attempt the author tries in this paper might be regarded as the first one with the aim to (re) discover the 'depth' in Japan in the information era and at the same time the pioneer work to try to (re) gain the 'depth' in the fields of studies on information society and on the relation(s) between information technologies and the meanings of life with 'depth' and with 'width' beyond the differentiation of things and persons at the surface level too.

Categories and Subject Descriptors

K.4.1 [**Ethics**]:

General Terms

Human Factors, Theory.

Keywords

Reductionism. Oneness. Japanese culture. *Seke*n. *Ba*.

Kitaro Nishida. Robots. Artificial oneness.

1. IS JAPAN REALLY A DECLINING COUNTRY?

What the author wants to think about in this article is to seek a potential newer direction for information studies, information ethics including roboethics in Japan. The author uses the terms 'a potential newer direction' in the sense that Japanese people of today have almost lost their inner 'compass' to be needed to map their future. In fact, Japan has never experienced economic growth measured by GDP growth for the last 20 years and its high-tech industries (e.g. electronics industry) have lost the leading position in the world in terms of the figures of sales and the ability to produce attractive products in the worldwide market.

In addition to these bad situations, Japan has been facing another crisis since March 2011, when Japanese people experienced the beginning of the multiple disasters, natural disasters and human-made disasters, earthquakes, Tsunamis and the serious accidents of the nuclear energy plants in Fukushima.

The following table shows that Japan's GDP growth in the recent years. As this table shows, GDP growth is very weak in Japan.

Table 1. Nominal GDP growth of developed countries

(Year)	1994	1995	1998	2002	2006	2010
Japan	4860	5349	3937	3991	4356	5510
USA	7309	7664	9089	10980	13858	14598
UK	1080	1181	1478	1621	2483	2295
Germany	2148	2523	2178	2007	2903	3304
France	1368	1572	1469	1452	2256	2565

(By US$ billion conversion)(Original data: [3])

And the following table shows us that Japan is evaluated negatively in terms of rank of 'world competitiveness.' The rank of Japan as one of the most highly developed industrialized countries is worse than Hong Kong, Singapore, Malaysia and Taiwan.

Table 2. IMD world competitiveness ranking

	2011	2012	2013	2014
1	Hong Kong	Hong Kong	USA	USA
2	USA	USA	Switzerland	Switzerland
3	Singapore	Switzerland	Hong Kong	Singapore
4	Sweden	Singapore	Sweden	Hong Kong
5	Switzerland	Sweden	Singapore	Sweden
6	Taiwan	Canada	Norway	Germany
7	Canada	Taiwan	Canada	Canada
8	Qatar	Norway	UAE	UAE
9	Australia	Germany	Germany	Denmark
10	Germany	Qatar	Qatar	Norway
Japan's rank	26	27	24	21
total	58	59	60	60

(Other ranking in 2014: Malaysia 12, Taiwan 13, China Mainland 23)(Original data: [7])

The following table shows the number of suicide people in Japan. This table shows the case in 2005 and we know that 'unemployed or jobless' is the factor leading to the most high suicide rate among others. The number of suicide people has been over 30,000 since 1998 and has continued to be the same almost for 20 years.

Table 3. The main causes of suicide in Japan in 2005

	Family-ownership business or independent business	Manager	Employee	Un-employed or jobless	Total
Family	116	15	279	447	1009
Health	313	52	595	2513	4087
Money	907	122	1093	1179	3436
Employ-ment	54	66	453	48	628
Without suicide notes	2372	380	5079	10751	21882
Total	3858	654	7893	15446	32325

(When we see the figures about 'unemployment or jobless,' we can understand that the problem of suicide reflects the total or plural situations in Japan's society.) (Original data: [17])

'Decreasing birthrate and aging of the population' (how to find useful actions to counteract the 'falling birthrate and ageing-shrinking population' problem) is another serious problem Japan is facing. Concerning this matter, Economic and Social Research Institute of Cabinet Office of Japan (Discussion Paper Series No.295 by ESRI, Economic and Social Research Institute of Cabinet Office) provides us with the findings making us face a very important and also serious phenomenon: young women in the age 20s and 30s with children (a child) feel *less happy* than those without children. (The research was done in Japan in 2012).

On the other hand, ONK (Office for national statistics) data of UK shows another useful suggestion on this matter. According to their data [20], the households with children tend to show affirmative attitudes toward the 'worthwhile' question than those without children, although the degree of feeling 'satisfaction' in life is not different among them.

This means that we might be able to have diverse views on the meanings of life which are not confined to personal emotions such as feeling of happiness and satisfaction. Without doubt, subjective feelings of happiness and satisfaction are among the most important index to evaluate the meanings of life, but we might widen the standards of evaluation in this point.

And as we will see in the next section, we know that Japanese people live in the cultural *Ba* with some sorts of 'depth': depth which might lead to broader meanings of life and the world at least at a latent level.

2. FROM WHERE DOES THIS CRISIS COME?

The figures of various tables seem to make us face the problems seriously: 'Is Japan really a declining country?'; 'From where does this crisis come from?'

Regarding these problems, the author's answer is very simple: the crisis comes from the fact that Japanese people lack the common perspectives (at least at the surface level) with which they can see

what is /has been happening in their life and society. This is because, the author thinks, their life has been divided into several contradictory pieces as the result of the 'import' of 'principle of competition' and the 'theories of information society' grounded on 'classical-computationalism' or 'reductionism' in the 1980s and 1990s.

It is very strange that Japanese people themselves are not aware of the 'true' situations of their society.

For example, in spite of the data showing Japan's declining tendency, we have a lot of different data showing the opposite tendency. In fact, the GDP growth of Japan per the person in age 15-64 (working age population) is almost twice that of the USA and higher than the major EU countries (in the period of 2000-2010). Japan's level of technology is still among the best in the recent years (2007-2009) in terms of the number of 'patent family' (a set of patents taken in various countries to protect a single invention, i.e. the number of the patent which shows the deduplicated number).

Table 4. The number of patent applications (2007-2009)

	Number of patent family	Share	Rank
Japan	59170	28.5%	1
USA	45308	21.8	2
Germany	30017	14.5	3
Republic of Korea	17533	8.4	4
France	10986	5.3	5
China	10431	5.0	6
Taiwan	9775	4.7	7
UK	8417	4.1	8
Canada	5501	2.7	9
Italy	5496	2.6	10
Netherlands	4631	2.2	11
Switzerland	3936	1.9	12

(Original data: [19])

After having examining these data, what we have to do is clearer; to find the reasons why the Japanese life has been divided into these contradictory pieces and how we can find the ways to combine these divided parts into a more integrated one.

One of the potential answers to the first part of this task is at least partly clear; Japanese people are/have been under 'reductionism' or 'techno-determinism' which is strongly related to 'information studies with a limited scope.'

In this case, 'limited scope' means that this kind of information studies and social information studies (studies on information society) are fundamentally based on the hypothesis of linearity or 'classical-symbolism'(or 'classical-computationalism').

We know that information studies or related research fields such as studies on AI (artificial intelligence) or robotics have experienced a kind of 'paradigmatic turn' in the 1980s or 1990s. And we know also: this turn means the emergence of 'connectionism' or 'DSA (dynamical systems approach)' as alternative models with the aims to overcome 'classical-symbolism' or 'classical-computationalism.' As we know, the discussions by Dreyfus [4], Brooks [2], Agre and Chapman [1], Winograd and Flores [25] and others since 1980s are the important ones to look for alternative models in this sense.

In spite of this new trend in the studies on information, AI and robotics in the other technologically developed countries, strangely enough, in Japan, it seems that the majority of scholars and authors in the fields of researches on information studies, social-information studies including information ethics are not aware of this 'paradigm shift.'

3. POTENTIAL DEPTH IN JAPANESE CULTURE AND SOCIETY

In the author's view, the data shown above might suggest us that Japanese people have lost the capacity to understand topics which need a broad range of sight including ontological perspectives on life (as we saw: 'what is happiness' might be dependent on plural ways of seeing the meanings of life). And it seems that this is nothing but the crisis Japanese people have been facing.

But on the other hand, according to the author's research data done in the previous 20 years or so with Japanese respondents, Japanese people share a set of common beliefs in or sympathies with certain types of traditional cultural and ethical views on this world and the meanings of life.

To put it another way, it seems that Japanese people still live in a (an alternative) *Ba* with some sort of cultural 'depth' which might be able to provide people with views to (re) find the meanings of the world in different ways than in the ways people have accustomed to regard 'normal': 'normal' but 'without depth' in the sense that our life can't be calculated by mere linear algebra.

Before examining the content of the data, we need the explanations on '*Ba*,' 'depth' or '*Ba* with depth.'

The author uses the term, 'depth,' because we can't see or experience this 'depth' by certain sorts of calculation or logics, i.e. calculation based on linearity or logics on the premise of linearity of this world.

And the author uses the term, '*Ba*'(place, locus), following Japanese cultural and philosophical tradition putting emphasis on the fact or the belief that at a deeper level the surface distinctions of the phenomena of our world, the distinctions between '*mono* (things, entities) and *koto*(words, expressions, experiences related to forms of narratives) ,' 'the subject and the object,' 'minds and

bodies,' 'the ways of understanding based on logical thinking combined with causality and the ways of understanding based on ethical views on things' would disappear.

To put it in a different way, this presupposes that there must be a (an alternative) place where these distinctions would disappear and the different cultural context would emerge. In this place (*Ba*) or the different cultural context, it is believed that the some sort of 'direct bonds among persons-things (*mono*), inner minds(*kokoro*)-outer events(*koto*), persons-persons, things(*mono*)-events(*koto*)' would be sensible through particular forms of expression or perception.

We need more explanation on this point.

In the author's view, one of the most important characteristics of Japanese traditional culture is its orientation to 'oneness' as the principle of the world or the index of values from which people see and evaluate the meanings of life or the world itself.

This oneness principle, *Ichinyo* in Japanese, is related to the view: even though at the surface level, *mono* (things), *koto*(events or words used to express *mono* and events), *hito*(persons) are divided into different entities or beings, but at a deeper level these divided or differentiated entities or beings would(could) be in the state of 'reciprocity,' interrelation or in the state of linkage.

In Japanese literature, we can find abundant examples of this orientation to oneness or *Ba* associated with oneness (oneness needs a place, *Ba* where this sort of un-differentiation or linkage at a deeper level would emerge) and shared sympathy with oneness and *Ba*.

In literature, this orientation to oneness or *Ba* related to oneness can be found together with 'mediated–indirect ways of expression of common/shared senses or emotions' which show another important aspect of Japanese culture too. The typical cases of 'mediated–indirect ways of expression of common/shared senses or emotions' can be seen in literature expressions like *Haiku* or some films such as Ozu's films.

In the author's interpretation this kind of 'mediated–indirect ways of expression of common/shared senses or emotions' are understood as 'restrained expression' or 'tendency to avoid abstract concepts in various aspects of life' or 'tendency to avoid straightforward emotional expressions' which is supposed to be the way to try to extract 'direct bonds among persons-things(*mono*), inner minds(*kokoro*)-outer events(*koto*), persons-persons, things(*mono*)-events(*koto*).' (With regard to *koto*, *mono*, see [10]).

It seems that without this kind of restriction the (re)finding of the links of *mono*, *koto* and *hito* would be difficult because in Japanese literature *hito*'s strong will or subjectivity is understood as the factor working negatively for surfacing of the latent links of *mono*, *koto*, *hito* which are believed to be in the state of 'passive synthesis' in the term by Edmund Husserl.

Daisetsu Suzuki [23], one of the well-known scholars on Zen-Buddhism, tries to turn our eyes to the importance of the interaction of *mono* and *hito* at a deep level.

According to Suzuki, one of the poems (*Haiku*) of Matsuo Basho (1644-1694) shows us about 'What kind of interaction is this?'

Yoku mire ba/ nazuna hanasaku/ kakine kana

(Look at there / a shepherd's purse here/blooming so quietly under this simple hedge)

Suzuki says that in contrast to Western eyes which are apart from the objects, Basho's eyes can't be separated from the bloom's presence, to be here.

In the author's interpretation, in somewhere between the object and the subject some sort of reciprocity might be considered to occur. In this somewhere (somewhere in-between) the difference of the viewer and the object might disappear and 'reciprocity' seems to appear.

In this case, 'reciprocity' means that the relation between the subject and the object is made possible through newly emerging form of observing and expression and this new form belongs both to the world of the subject and to the world of the object. In this sense, in this case we might say that a new locus (*Ba*) would appear as a place: the place where this reciprocity is considered to emerge. And this place or *Ba* is just what Kitaro Nishida, Japanese philosopher, (and others) has tried to intuitively and logically understand.

Kitaro Nishida is one of the most well-known philosophers in Japan who tried to turn the eyes of Japanese people to this kind of experiences or phenomena happening in *Ba*.

According to his discussions in his 'Tetsugaku gairon,' the true reality comes from the direction indicating the place (*Ba*) where a copula works [18].

He says that the judgment is not the work of the subject which is (in many cases) considered to exist as the first entity and extract the accompanying work of the predicate. He says that this is a misunderstood thought or at least a culturally biased thought. And also the opposite is due to the misunderstood thought or the culturally biased thought: the predicate or something universal is the first and the subject or something individual is the second. He says that the true reality emerges through the linking work of a copula which links the subject (the subject-related matters) and the object (the object-related matters). And this joining work makes a 'Gesamtvorstellung' emerge. ('Gesamtvorstellung' is Wilhelm Wundt's term and refers to comprehensive representation or entire representation.)

This linking work is happening somewhere beyond the distinction of the inner world of the subject and also the outer world of the object. This is happening in somewhere called be 'in-between'('*Aida*' in Japanese) or *Ba* related to oneness.

Stanford Encyclopedia of Philosophy adds the additional explanation to this 'Gesamtvorstellung,' pointing out that this is an idea leading to denial of differentiation of the reality into the subject and the object as a primordial form of reality [22].

We know that Nishida's idea of 'pure experience' or 'The Good' is understood as a situation of realization of "the fundamental form of reality" or "a higher unity" [22]. And we know that some other authors or scholars in Japan share a similar interest in 'the fundamental form of reality,' 'a higher unity' or *Ba* where this kind of reality or unity is made possible.

Yujiro Nakamura [16], a Japanese philosopher who attempts to combine Japanese traditional thoughts with Western modern thoughts, suggests that Kitaro Nishida tried to regain the meanings of beings based on *Mu* (nothingness) or 'predicative substratum' ('substrata') which is in contrast with subjective substratum ('substrata'). In this sense, *Mu* is understood not as mere emptiness but as the source of beings (*Yu*) on which articulations of beings are based.

According to Nakamura, oneness of *Mu* and *Yu*, or oneness of subjects and objects, oneness of events (*koto*) and words

(*koto=gen*) needs *Ba* (or *Bamen* = place, field) where a 'coming together' of subjects and objects, events (*koto*) and words (*koto=gen*) is possible. This *Ba* or *Bamen* includes, as Nakamura stresses while quoting the work of Motoki Tokieda, a Japanese linguist[24], things, scenes, subject's attitudes, subject's feelings, and subject's emotions[16].

These concepts, 'pure experience' or 'The Good,' 'the fundamental form of reality,' 'a higher unity' or *Ba* would be nonsense or meaningless from the world views based on modernized and rationalized Western culture(s) and also are regarded as 'worthless' from Japanese views putting emphasis on such values as profits, competition , efficiency and so on.

But if we try to pay attentions on presuppositions of these rationalized and undoubted views, we might find 'from where do these views come?' We might be able to say that these views consist of undoubted acceptance of differentiation of this world into individualized or isolated entities or beings of *mono*, *koto*, *hito*, mind and body. And we know the miscarried or discouraged plans to produce autonomous robots or AI based on 'classical computationalism' or 'classical symbolism' are the ones which fundamentally reflect this kind of differentiated word views.

4. REDISCOVERY OF JAPANESE *BA* WITH DEPTH

The orientation to oneness or *Ba*-related meanings associated with oneness is not confined to literature. This is what the author himself found through his own researches about *Seken*-related meanings or phenomena which seem to lie in Japanese minds of today.

Seken is, according to the author's interpretation, the realm of society or world consisting of meanings which seem to come from some sort of oneness or related situations.

Seken is considered to be the place where people want to understand or talk about the meanings of social phenomena, accidents, disasters in the ways reflecting shared common beliefs in the 'true meanings' of life and the world in the sense that minds(*hito*'s minds) and phenomena(*koto* as social and cultural events) are not differentiated at a deeper level.

The following is the citation (translated into English by the author) of an editorial of Asahi Shimbun to be appeared on February 5, 1995 after the Hanshin Daishinsai(Great Hanshin Earthquake).

The editorial says:

'I do not want to think that even the natural disaster of this time was something beyond the human understanding. In huddled in the natural side and being difficult to see, but if you would have looked at with eyes with a long-range scope, they might have been predicted. However, for our entire society keeping shutting eyes to the risks of the nature, there must have been a blind spot in our technological civilization. We have been too eager to live a life with material convenience and comfort and this must have led to this regret showing us the lesson that the civilization itself might be shaken from the ground when we neglect to prepare for the risk.'

This is a typical attitude toward natural disasters or other catastrophic phenomena found in Japan. Fundamentally this kind of attitudes derive from a traditional culture consisting of traditions of Buddhism, Confucianism, Shinto(Japan's indigenous religion), the memories of experiences of the disasters and the wars in the past history or the senses to the linkages of *mono*, *koto* and *hito* mentioned above. The attitudes or the meanings, which the author have called '*Seken*-related meanings' or '*Seken*-related attitudes (toward the world)' in my previous papers, are the ones which the author has found through his previous quantitative and qualitative researches done in Japan and in some other countries in the last 20 years.

The etymology of the term *Seken* derives from the Sanskrit word 'loca.' Originally the meaning of *Seken* comes from *Se* (time or transient situations of this world/life) and *Ken* (in-between or locus), i.e., the transient *Ba* consisting of transient human activities and the place where these activities are done[8].

Although not a few Japanese scholars and authors have been talking about *Seken*, it is very rare that their discussions on *Seken* are based on the quantitative researches or on the combination of the quantitative and qualitative researches. In fact, the author's or his research group's researches are the first ones which have succeeded in extracting the link of the meanings coming from *Seken*-related experiences or cultural traditions.

Through the research done in 1981 in the areas of Tohoku, the area with a long history of severe damage by the destructive Tsunamis (tidal waves) for decades or even centuries, the author and his colleagues found an important fact: *Seken*-related meanings and attitudes are still active in people's minds.

The author has tried to continue to do the similar researches in Japan since 1981. And as the following table shows, the findings gained through 1981 research have been repeatedly found.

Table 5. Sympathy for *Seken*-related meanings

	1995G	2000G	2011G	2013JS1	2014G
Distance from nature	**73.6%**	-	**78.0**	**67.8**	**71.2**
Honest poverty	**83.7**	**81.5**	**87.0**	**81.5**	**80.4**
Destiny	**84.4**	**79.0**	**82.4**	**69.1**	**77.5**
Denial of natural science	**88.5**	**88.3**	**88.2**	**87.4**	**81.8**
Criticism of selfishness	**85.5**	**88.3**	**80.3**	**59.6**	**76.8**
Powerlessness	**71.9**	**64.8**	**77.8**	**53.3**	**72.7**
Superficial cheerfulness	**73.3**	**65.6**	**72.7**	-	**70.0**
Belief in kindness	-	**68.1**	**74.3**	**78.9**	**66.5**
Scourge from heaven	**62.7**	49.5	-	-	-
Warnings from heaven	-	-	**60.2**	26.1	**59.0**

1) Table 5 shows the percentages of the respondents who said 'agree or somewhat agree' to the statements of *Seken*-related meanings. These statements are: "Within our modern lifestyles, people have become too distant from nature"(Distance from

nature); "People will become corrupt if they become too rich"(Honest poverty); "People have a certain destiny, no matter what form it takes"(Destiny); "In our world, there are many things that cannot be explained by science"(Denial of natural science); "There are too many people in developed countries (or Japan) today who are concerned only with themselves" (Criticism of selfishness); "In today's world, people are helpless if they are (individually) left to themselves" (Powerlessness); "In today's world, what seems cheerful and enjoyable is really only superficial" (Superficial cheerfulness); "Doing your best for other people is good for you" (Belief in kindness); "The frequent occurrence of natural disasters is due to a scourge from heaven" (Scourge from heaven); "Occurrences of huge and disastrous natural disasters can be interpreted as warnings from heaven to people"(Warnings from heaven). 2) Figures in **bold type** indicate the items to which over 50% respondents showed affirmative answers.

(Note: the researches shown in Table 5.

'2013JS1': the respondents are 379 university students from 2 universities located in Tokyo Metropolitan Area. '2014G': the research done in 2014 from August 29 to September 2, 2014. The 729 respondents (Internet users living in Fukushima, Miyagi and Iwate Prefectures) were selected by a research company in Japan. This survey was designed as quota sampling, and ratios of gender and age were quoted from the official statistical report of the Japanese government about the Internet users in 2010 in Japan.

Concerning the concrete explanation on the other researches, see: [11, 12.14].)

In addition to the findings shown in the table mentioned above which suggest us the presence of a set of '*Seken*-related meanings' in Japanese minds of today, the other important findings from these researches are: the place(*Ba*) including *Seken*-related meanings is also the place(*Ba*) where people find the relations between different sorts of meanings. For example, the following table shows that there are the relations between 'the meanings of technologies and technological products such as robots' and 'the meanings of life at a deep level' in the case of 2014G research.

By examining 2014G research data, we could find that '*Seken*-related factor 2'(one of the factors gained through factor analysis, with the methods of principal component analysis and Varimax rotation, on the data of '*Seken*-related meanings') has statistically significant correlations with other views on different horizons in terms of meanings. *Seken*-related factor 2 is a factor including 'orientation to good human relations through sincerity,' '(sympathy for) warnings from heaven' and '(sympathy for) destiny.'

Table 6. The relations between '*Seken*-related factor 2' and other important meanings of life and technology(Data: 2014G)

Seken-related factor 2	'interest in environmental problems(0.352**,i.e. correlation coefficients=0.352 with statistical significance at the level of 0.01),' 'interest in political problems (0.182**),' 'interest in activation of local areas(0.404**),' 'concerns for ethical problems of use of robots in the everyday life (0.519**),' 'sympathy for robots as artificial life (0.245**),' '(sympathy for)sacrifice and "being beautiful through transience (0.325**)",' '(sympathy for)victims(0.425**)'

In the author's interpretation, the 'import' of 'principle of competition' and the 'theories of information society' grounded on 'classical-computationalism' or 'reductionism' in the 1980s and 1990s has made this kind of 'depth' invisible in various areas of Japanese society, because these meanings are not the ones which can be gained through some sort of calculation.

The following table shows that Japanese *Ba* includes various meanings, which seem to derive from different horizons of meanings and that these meanings come together in this *Ba*.

Table 7. Robot-related factors and the meanings of life at a deep level（Data:2014G）

	Interest in political problems	Interest in global environment problems	Interest in activation of local areas	Robot 1 (criticism for robot's use)	Robot 2 (sympathy for robots)
Sympathy for victims	.170**	.246**	.319**	.262**	.184**
Sympathy for 'sacrifice' and 'being beautiful through transience'	.321**	.343**	.442**	.519**	.245**

1)**=p<0.01, *=p<0.05 2) 'Robot 1,' and 'Robot 2' refer to the factors gained through factor analysis on robot-related views. 3) *Sympathy for victims*(the factor gained through factor analysis on the data concerning the meanings of 'sacrifice,' 'beauty through transience,' 'lonely death' and so on) includes the sympathy for such matters: "I can imagine clearly the figures of the victims or their familes when I see the flowers for lament or sorrow at the places of traffic accidents or other accidents" and the like. *Sympathy for 'sacrifice' and 'being beautiful through transience'* includes the sympathy for such matters: "I sometimes feel that I have to think more deeply about the important meanings of life when I hear the stories of persons who saved others at the cost of their own life in natural disasters and similar crises."; "I can

sometimes feel that the fireworks or the glow of a firefly in the summer are beautiful because they are transient or short-lived."

5. ARTIFICIAL ONENESS OF THINGS AND HUMANS

It seems that Japanese culture is characterized by orientation to oneness of things-humans, emotions-judgment, or 'direct bonds among persons-things (*mono*), inner minds (*kokoro*)-outer events (*koto*), persons-persons, things (*mono*)-events(*koto*)' [9,24].

Fukada[4] shows us a very interesting suggestion about this. According to Fukada's paper which focuses on the 'undifferentiated meanings' of certain types of adjectives, Japanese language includes abundant adjectives which can be interpreted to show the situations of the objects or the phenomena and at the same time can be interpreted to show the situations of the person(who uses the adjectives).

For example, in the case of '*isogashii toori*'(busy street), '*isogashii*' shows the situation of the street or/and the situation of the subject(or the subjects) who takes that street.

And it seems that this is the situation itself noted carefully by Kitaro Nishida, Japanese philosopher.

If we can join this discussion about adjectives in Japanese language with the other undifferentiated meanings, we might be able to understand 'From where do the senses of "oneness" in Japanese culture come?' at least partly.

And we might say that the relations among the things, persons, situations on different horizons mentioned above (e.g. the relation between 'interest in politics' and 'sympathy for sacrifice') come from this kind of undifferentiated meanings (these are related to undifferentiated experiences at the same time).

We might have an impression that this kind of 'oneness' would be meaningful only in the areas of poetic imagination or emotions related to literature, in particular in Japanese literature or the phenomena expressed through Japanese language. But according to the findings gained in the fields of studies on neuroscience, cognitive science, AI and robotics which are based on 'scientific' methodologies, it seems that we are now experiencing a lot of experiences similar to this kind of 'oneness' in the sense: we experience the situations where *mono*, *koto* and *hito* seem to be connected in the ways beyond the subject-object dichotomy.

It seem that we are involved in this kind of oneness in everyday activities, although the scientists and researchers in these fields seldom use the term, 'oneness' and we are usually without being awareness of the fact that various important meanings come from this oneness.

Sasaki et al. [21] show us a very interesting example of oneness, but a kind of oneness which might be called 'artificial oneness,' i.e. oneness which can be found through artificial situations made possible by technological products or technological devices.

They asked the subjects (the experiment participants) to move their hands upward or downward (drag a centrally located dot towards a cued area, which was either in the upper or lower portion of the screen) just after watching an emotional image on a touch screen. What Sasaki et al. found is that the rate of emotional valence of the images is statistically associated with the upward or the downward hand movements. They found that upward hand movements are associated with positive values of the stimuli and the downward hand movements are associated with negative evaluations.

We might think that these phenomena are just limited within human experiences and the subjective situations and things themselves are apart from this kind of integration. But even so, it seems that this kind of integration might influence the planning or designing of things including technological products.

Even if the term 'oneness' is not clearly used, it seems that many scholars of the neuroscience and cognitive science virtually use the similar frames which might be related to oneness in some ways.

In the work of Guterstam et al. [6] the terms 'visuotactile-proprioceptive,' 'the integration of multisensory body-related signals' and 'the dynamic integration of signals from different sensory modalities' are used. Their experimental work is a kind of similar researches on 'rubber hand illusion.' Instead of applying brush strokes to the rubber hand, they continued to stroke (with a brush) to the empty space where the rubber hand is supposed to be located. Like the other similar experiments, after the experimenter has stroked the space (or the invisible -illusionary entity) in synchrony with their real hand (which is hidden from view behind a screen) for a while, most participants started to feel 'transferring of the sensation of touch from the real hand to the region of empty space where they see the paintbrush moves,' i.e. rubber hand illusion.

The Japanese traditional term 'oneness' of *mono* and *hito* is usually applied to the case, 'primordial bonds between *mono* and *hito*' and in the case of this experiment we can say that the artificial integration of *mono* and *hito* can emerge through some sorts of artificial settings or environments. The important point about this is that these types of artificial oneness can't be calculated by linear calculation methods in many cases in the sense that we don't know why this kind of phenomena appear in a logical way and therefore we can't calculate the presuppositions of these phenomena.

In the case of robotics the emergence of 'entrainment' (integration of body movements of different entities like robots, parts of robots' bodies or parts of robots) might be considered to be some sorts of oneness in the sense that the cases of emergence occur in *Ba* beyond the distinctions of things and humans, things and events. And as we know, the phenomena related to this *Ba* or oneness come from the work or the function of artificial neurons, CPG(central pattern generators) or the body schema including legs and arms grounded on DSA(dynamical systems approach).

In the author's view, this might be considered to be another case of oneness including some kind of autonomous movements called 'entrainment'.

6. CONCLUSIVE REMARKS

The main topic in this paper might be summarized: we have to see 'depth' in our culture(s) and in our technological environments.

As we have directly and indirectly discussed, our culture and our technological environments seem to be considered to have various meanings at least at a latent level. But as we have seen, our scope for understanding the meanings of life and the world might have been affected by strong influence deriving from the imported 'principle of competition' and the 'theories of information society' grounded on 'classical-computationalism' or 'reductionism.' This seems to be the factors affecting the crisis mentioned above in Japan. In this paper, we have focused on the problems in Japan, but as we know, the crisis might not be confined to Japan in the sense that loss of awareness of the

meanings with 'depth' comes from 'reductionism,' 'undoubted presuppositions of linearity' or 'the subject and object dichotomy as undoubted world views leading to forgetfulness of oneness.' So the aim of this paper might have the depth or width with which we could see 'What is happening in the world in the information era ?'

7. REFERENCES

[1] Agre, P. E. and Chapman, D. 1990. What Are Plans for? In Pattie Maes (Ed.), *Designing Autonomous Agents*: *Theory and Practice from Biology to Engineering and Back*. The MIT Press, Cambridge, MA.

[2] Brooks, R.1986. A Robust Layered Control System for a Mobile Robot. *IEEE Journal of Robotics and Automation*, 2 (1), 1986, 1-26.

[3] Cabinet Office, Government of Japan.2014. Retrieved November 25, 2014 from http://www.esri.cao.go.jp/jp/sna/menu.html/

[4] Dreyfus, H.1972.*What Computers Can't Do*: *A Critique of Artificial Reason*. Harper and Row.

[5] Fukada, Chie. 2004.Mibunkana imi no bunnka(differentiation of undifferentiated meanings). *Papers in linguistic science* (2004), 10: 117-147(University of Kyoto).

[6] Guterstam, A., Gentile, G. and Ehrsson, H.2013. The Invisible Hand Illusion: Multisensory Integration Leads to the Embodiment of a Discrete Volume of Empty Space. *Journal of Cognitive Neuroscience*, July 2013, Vol. 25, No. 7, 1078-1099.

[7] IMD World Competitiveness Yearbook Ranking.2014. Retrieved May 22, 2014 from http://www.imd.org/news/2014-World-Competitiveness.cfm/

[8] Inoue, Tadashi.1977. *'Seken-tei' no kouzou*(the structure of the reputations in *Seken*). Nihonsyuppankyoukai, Tokyo.

[9] Izutsu, Toshihiko. 2001. *Ishiki to honshitsu*. Iwanami, Tokyo.

[10] Kimura, Bin.1994. *Kokoro no byourigaku wo kangaeru*. Iwanami,Tokyo.

[11] Nakada, Makoto. 2011a. Japanese views on privacy and robots in Japanese ethical *Ba* (place). In Jeremy Mauger (Ed.) *CEPE2011*, published by INSEIT, 208-202.

[12] Nakada, Makoto.2011b. Ethical and critical views on studies on robots and roboethics. In Michael Decker and Mathias Gutmann(Eds.)*Robo-and Informationethics*: *Some Fundamentals*. Lit Verlarg, Berlin, 157-186.

[13] Nakada, Makoto.2011c. Ethical and Critical Analysis of the Meanings of 'Autonomy' of Robot Technology in the West and Japan: Themes Underlying Discussions of the Ethical Aspects of Robotics and Human-Robot Interaction. In T. Kimura, M. Nakada, K. Suzuki, Y. Sankai(Eds.) *Cybernics Technical Reports* (Special Issue on Roboethics), 61-91.

[14] Nakada, Makoto. 2012. Robots and Privacy in Japanese, Thai and Chinese cultures: Discussions on Robots and Privacy as Topics of Intercultural Information Ethics in 'Far East. In Fay Sudweeks, Herbert Hrachovec and Charles Ess(Eds.) *Cultural Attitudes towards Technology and Communication 2012*. Murdoch: Murdoch University, 200-215.

[15] Nakada, M. and Capurro, R.2013. An intercultural dialogue on roboethics. In Nakada.M. and Capurro, R.(Eds.) *The Quest for Information Ethics and Roboethics in East and West*, vol.1,13-22.

[16] Nakamura, Yujiro. 2001. *Nishida Kitaro I*. Iwanami, Tokyo.

[17] National Policy Agency of Japan.2014. Retrieved November 30, 2014 from http://www.npa.go.jp/safetylife/seianki/jisatsu /H16/H16_jisatunogaiyou.pdf/

[18] Nishida, Kitaro. 1966. Tetsugaku gairon. *Nishidakitaro Zennsyuu 15*. Iwanami, Tokyo.

[19] Nistep repository, Japanese science and technology indicators 2014.2014. Retrieved 1 January,2015 from http://data.nistep.go.jp/dspace/handle/11035/2490/

[20] ONK (Office for national statistics) data of UK.2011. Analysis of Experimental Subjective Well-being Data from the Annual Population Survey, April to September 2011.Retrieved January 28,2014 from http://www.ons.gov.uk/ons/rel/wellbeing/measuring-subjective-wellbeing-in-the-uk/analysis-of-experimental-subjective-well-being-data-from-the-annual-population-survey--april---september-2011/report-april-to-september-2011.html/

[21] Sasaki, K., Yamada, Y. and Miura, K. 2015. Post-determined emotion: motor action retrospectively modulates emotional valence of visual images. *Proceedings of the Royal Society B: Biological Sciences*, 282: 20140690/

[22] Stanford Encyclopedia of Philosophy.2012. Retrieved August 22, 2014 from http://plato.stanford.edu/entries/nishida-kitaro/

[23] Suzuki, D., Fromm, E. and De Martino, R. 1960. *Zen to seishin bunseki*(Zen Buddhism and Psychoanalysis). Sougensya, Tokyo.

[24] Tokieda, Motoki. 2008. *Kokugogaku genron*. Iwanami, Tokyo.

[25] Winograd, T. and Flores, F. 1986. *Understanding Computers and Cognition*: *A New Foundation for Design*. Ablex, New Jersey.

False Friends and False Coinage: A tool for navigating the ethics of sociable robots

Alexis Elder
Philosophy, Southern Connecticut State University
501 Crescent Street
New Haven, CT 06515 USA
+12033926794
alexis.elder@gmail.com

ABSTRACT

In this paper, an analogy that Aristotle drew between false friends and false coinage is leveraged to identify ethically important features of cases involving so-called sociable robots. The use of such robots to care for the elderly and disabled poses both benefits and costs. Although a uniform verdict on the ethical use of these robots is unlikely to be forthcoming, owing to the importance of context and wide array of variables that can influence assessment of a situation, progress can be made by using analogies from other domains. Such analogies can help identify relevant features of a given situation, in order to better evaluate the costs and benefits to patients, caregivers, and designers, thereby facilitating appropriately context-sensitive judgments.

Categories and Subject Descriptors

K.4.2 [**Computers and Society**]: Social Issues – *Assistive technologies for persons with disabilities, employment*.

General Terms

Design, Human Factors.

Keywords

Robots; friendship; geriatric care; virtue ethics.

1. INTRODUCTION

I begin by introducing the technology under consideration: so-called sociable robots. I survey some of the advantages of using this technology, as well as some of the potential drawbacks. I introduce my solution, and then compare it to other extant accounts of the ethical benefits and perils of the technology. I conclude by showing how the analogy can be used to make sense of several complex considerations likely to arise in the real-life implementation of these robots.

2. SOCIABLE ROBOTS?

One major area where robotics shows great promise is in the field of healthcare, including a number of applications involving care of the elderly. In this paper, most of my examples will involve the use of robots in geriatric care, but much of what I say will also be applicable to other applications in healthcare fields.

Robots can be used to address a number of standing concerns in geriatric care. They can help offset staffing shortages [15], which are of particular concern in places where the elderly constitute an increasingly large percentage of the population (a consequence of falling birthrates), as is the case in many parts of Asia, Europe, and the Americas. Robots can increase the autonomy of geriatric patients, helping them to perform tasks they would otherwise be unable to manage (physically or mentally) on their own [4]. Robots, at least when properly designed and built, can provide a consistent quality of care that would be likely to constitute a significant improvement over the notorious inconsistencies of human geriatric caregivers [18]. Robots can perform as well as human caregivers at some tasks. In fact, robots may be preferable in circumstances such as assisting with bathing, and using the toilet. Seniors who value privacy and dignity may prefer to use non-human assistance in such situations rather than exposing themselves to even the best human aides [11]. Robots can be customized to address the particular needs of different patients. And they can be a constant presence for both housebound and institutionalized seniors where it would be cost-prohibitive to provide round-the-clock human assistance.

They thus show promise on a number of fronts.

The application I will focus on is the use of robots to meet seniors' social needs. This may overlap with a number of the implementations sketched above, but has also proved a compelling enough area that some robots have been developed specifically for this purpose.

Loneliness and lack of social connection are widely recognized problems in geriatric care. So-called sociable robots are designed to address these issues. One of the most famous is Paro, a robotic baby seal, first introduced in Japan and since adopted by eldercare facilities throughout Europe and the United States. Paro has even been approved as a medical device by the FDA (Food and Drug Administration, a US regulatory agency) [10]. Paro and other sociable robots are typically designed with highly anthropomorphic features, big "cute" eyes, and facial or postural expressiveness, which elicit strong emotional reactions. They are often designed to be highly interactive and responsive via auditory, visual, or tactile channels. Some respond to spoken words, others to gestures. Paro differentiates between light touch, such as stroking, and hard contact such as striking. It responds by adapting its activity to avoid provoking reactions like striking, which may indicate that something it has done has been upsetting to patients, and increasing the frequency of behavior that results in

petting and scratching. They can thus tailor their behavior to reflect the preferences of particular users.

It is an open question whether seniors' needs will best be addressed by dedicated, specialized sociable robots, or whether these features will eventually be implemented in robots designed to serve other caregiving purposes as well. So what I will argue here, although drawn primarily from the current generation of sociable robots, may have implications for geriatric care robots more generally. Fortunately, Paro and other sociable robots have been in use for long enough that we have some data on their impact on geriatric patients. This allows us to say with some confidence what some of the issues of current and future iterations of sociable robots may be.

Robots have been shown to reduce patient loneliness at rates comparable to those achieved by animal-assisted therapy, in side-by-side comparisons [2]. Furthermore, this reduction in loneliness can offer at least one significant advantage over animal-assisted therapy: it does not introduce the sanitary concerns of using live animals around patients who may suffer from weakened immune systems [7, 16]. Patients' testimony supports their utility. A 74-year-old patient in a nursing home who was exposed to a sociable robot reported, "The first time, I didn't like playing with the robot because I was depressed. After I had played with the robot several times, I felt good." A 68-year-old patient said, "I do not think about anything while playing with the pet-type robot. It heals my mind" [7].

We can distinguish between benefits that we might consider intrinsic to something, and those are grounded in its being instrumental in producing some further, independent benefit. Patients who find enjoyment and cessation of loneliness in interacting with sociable robots enjoy what we might think of as an intrinsic benefit. But there are significant instrumental benefits to these interactions, as well.

Reducing loneliness is an important health issue. Loneliness can, of course, be a component of general quality-of-life concerns: all things being equal, it seems likely that most of us would prefer to avoid the lonely life, even if it were not associated with any additional costs. Loneliness, however, is also of concern in geriatric care because it can have a significant impact on the patient's health in a variety of ways. "An extensive social network seems to protect against dementia", report Fratiglioni et al. Even infrequent social contacts can help reduce occurrence of dementia "if such contacts were experienced as satisfying" [5]. Risk of Alzheimer's disease turns out to be sensitive to "perceived isolation", or loneliness, with risk "more than doubled in lonely persons compared with persons who were not lonely" [21]. Wada & Shibata observed the results of introducing Paro seal robots to elderly residents of a care house, and found that "urinary tests showed that the reactions of the subjects' vital organs to stress were improved after the introduction of the robots" [19]. Tamura et al reported that patients with severe dementia "recognized that AIBO [the sociable robot they used] was a robot. However, once we dressed AIBO, the patients perceived AIBO as either a dog or a baby. Nevertheless, the presentation of AIBO resulted in positive outcomes for the severe dementia patients, including increased communication between the patients and AIBO" [16]. These results show that sociable robots can provide significant instrumental benefits to geriatric patients.

There are other potential benefits to using sociable robots. As most people who have been involved in caring for elderly friends and relatives can attest, such work can be emotionally and psychologically taxing. Many cognitive and volitional conditions common amongst the elderly can contribute to mood swings, depression, tantrums, and outbursts, often compounded when caregivers have longstanding, complex, and emotionally laden histories with their charges, as when adult children care for elderly parents. Sociable robots can withstand and adapt to difficult moods and personalities without the emotional toll often suffered by family members and caretakers of patients with dementia, depression, and other challenging conditions. In this respect, sociable robots may be preferable because they can be deployed without exposing human beings to some of the costs of eldercare.

In addition, they can help address feminist concerns about the "care burden" of (predominantly women's) caretaking labor [9]. And although it may seem potentially problematic to use cutting-edge technology to care for those patients least likely to have adapted to the rapidly-changing technological landscape, the problems posed are often in principle similar to general design challenges for the general public. Tradeoffs between transparency and usability for nontechnical users, for example, are familiar issues for designers, not a special problem for those focusing on elderly users.

3. ETHICAL PROBLEMS OF SOCIABLE ROBOTS

However, the benefits listed above should not lead us to make an unqualified endorsement of their use. Sociable robots pose significant ethical challenges, as well. Some worry that the use of robots for geriatric care, particularly those parts that seem most emotionally and ethically important (such as addressing our need for social contact) ends up devaluing (human) care and reinforcing "broader social attitudes towards older persons" as unworthy of our attention and effort [15]. Because they elicit powerful emotional responses from users, who respond to them as if they were genuine agents, they are potentially deceptive [6]. We might worry that the use of simple, uncomplicated, appealing but quite "shallow" sociable robots provide short-term cessation of loneliness but ultimately decrease one's capacity for genuine connection in actual messy, complex, demanding human relationships [17]. If the use of sociable robots becomes widespread, we might worry that this will increase *actual* loneliness of patients (as opposed to perceived loneliness). That is, if it becomes common to use robots to care for the elderly, this may contribute to the isolation of seniors, insofar as it leads us to substitute robot relationships for human ones [12]. These different concerns can be mitigated in different ways, by technological fixes, policy changes, and responsible exercise of individual choice. As I will discuss in the section on Comparing Aristotle's Analogy to the Competition, some of these concerns may be less frightening than they initially appear. But more broadly, there seems to be a widespread if somewhat inchoate intuition that it is problematic to use robots to provide for our social needs. At least, it seems problematic given that our current crop of robots seems pretty far from the sort of thing one could plausibly consider a genuine moral agent. While there may remain the possibility that a robotic agent such as Data from *Star Trek: The Next Generation* could be a satisfactory companion, today's robots, and those we can expect in the near future, are simple and deterministic enough that few would consider them to be agents in any rich or interesting sense. Their use to satisfy social needs thus strikes many as problematic, although the challenge is to explain why this is so.

In what follows, I draw on the Aristotelian tradition to make sense of this intuition. Aristotle's eudaemonist theory of virtue strikes

many as a promising way to approach ethical questions involving emerging technologies. It invites us to consider what constitutes a good, flourishing life, and the impact of both large and small actions and influences on our ability to live such a life. [3, 18] In addition, Aristotle offers an exceptionally detailed and comprehensive account of the nature and value of friendships and close-knit personal relationships, which can be used to find guidance when confronted with complex social issues such as those posed by the introduction of sociable robots. I draw primarily on Aristotle's *Nicomachean Ethics,* in which his most mature ethical theory is presented, to make sense of his account of sociality. Two out of the ten books of the *Nicomachean Ethics* are devoted to friendship, reflecting the centrality he ascribes to friendship for the good life for a human being.

At the start of the first of these two books*,* Aristotle argues that friendship is valuable intrinsically, in its own right, not merely as a means to other ends: "no one would choose to live without friends even if he had all the other goods" [1]. Thus, the fact that sociable robots present significant instrumental benefits (as they surely do) could not, even in principle, justify them as meeting our social *needs* in the richer sense of providing us with the ingredients for a good life. That is, we should not expect the instrumental value of sociable robots to completely justify them in their capacity as companions. If they are to be justified, it will have to be by other means.

A thought experiment can be used to show that, even were one to develop a robot with which interactions were indistinguishable, to the patient, from those of a human companion, such a creation would not suffice to provide the full goods of friendship.

Imagine that you are given a choice between two possible lives. You know at the time of the choice how the lives will differ. But once you make your choice and begin your chosen life, you will forget ever having been given the choice – it will be as though this is the way things have always been. In one option, the people you consider your closest friends are actually paid actors, although if you were to choose this life, once it commenced you never learn of their illusory nature, and could never observe it from their actions. These friend-facsimiles would not use the appearance of friendship to exploit you. They would not betray your confidence, or use your trust to take advantage of you, but they would also not care for you or find pleasure in interacting with you. Call this the Truman Show option. In the other world, your closest friends are exactly as they appear to you to be. Call this the Genuine option.

It seems apparent that, given the choice, most of us would prefer the Genuine over the Truman Show life. This shows at least two things. First, that we value more than appearances in friendship, because one is preferable to the other and yet both provide the appearance of friendship. Second, that the Genuine option is choice worthy for non-instrumental reasons. In Truman Show, one gets all the instrumental benefits of friendship, because the friends outwardly act and provide all the same external goods as in Genuine. Furthermore, they cause none of the harms ordinarily associated with "false friends". And yet it is less attractive than Genuine. So whatever is good about Genuine cannot be the instrumental benefits and costs. Our preferences are not based on differences in experiential quality, nor in the bad outcome of having actors as "friends". This thought experiment shows that, the most choice worthy lives involve genuine friendships, which involve reciprocal caring of genuine agents capable of care. Such relationships must be intrinsically valuable.

So when we assess the value of sociable robots, it is not sufficient to show that patients accept them as companions to show that they provide us with the goods of sociality. (Recall that dementia patients presented with an AIBO thought it was a baby.) If the patients are unaware of the nature of the things from which they derive social satisfaction, they may not be in possession of the good they take themselves to have, even when they can be observed to reap the instrumental benefits of sociality. Unless the robots themselves are the sorts of things we can conceive of as agents who value us and take pleasure in our company, as the genuine friends do in Genuine, we run the risk of selling patients a false bill of goods, leaving them worse off than they think they are. This seems a high ethical cost of using sociable robots, even though it is not applicable in every case.

The powerful benefits and seemingly intractable costs of sociable robots identified lead me to believe that we should not expect a simple policy decision on whether or not their use is ethical. Rather, their implementation in eldercare involves a range of thorny issues. I agree with Vallor that in such matters, we should not look for a uniform principle to guide us. [18] Details of context matter, and identifying relevant features of context is a nontrivial task. [8] Most real-life contexts in which we must make ethical decisions involve tradeoffs between competing goods (e.g. between physical health and autonomy), and it is no different for decisions about deploying sociable robots. Complicating matters further, the varying cognitive, emotional, and volitional abilities of geriatric patients may influence what counts as a good choice for a given subject.

So what, then, should we look for, if not a uniform principle that tells us what to do? Ideally, we want to make wise and careful decisions that are appropriately sensitive to context, but we may need help to get there. An analogy may serve as a roadmap to help take stock of the territory, even if it cannot tell us exactly where to go. A good tool can help people think through particular cases and identify important costs and benefits in order to make intelligent tradeoffs.

The issues surveyed so far suggest that an appropriate tool for navigating ethical issues will incorporate the following considerations. As we have seen, some patients recognize that they are interacting with robots, while others, owing to various kinds of cognitive disorders, do not. We might characterize this as the difference between enchantment and deception. In enchantment, one recognizes that the thing with which one is interacting is not a person, but because features of the object appeal to one's emotional patterns of response (such as big cute "eyes", high-pitched vocalization, behavior modification in response to voice, touch, or facial expression), one goes ahead and interacts as if the item were a person. In robust deception, these or other features are sufficient to convince the individual that the object is a person (as with the dementia patients who thought AIBO was a baby).

In evaluating the wellbeing of patients, a broad and rich conception of wellbeing should be utilized. It should not be limited to thin, easily-operationalized concepts like subjective pleasure or lowered incidence of dementia, but should also include whether these patients are living good lives of the sort we consider choice worthy; whether their desires are really being met or only appear to be satisfied. Instead, something more eudaimonist, which includes both physical and psychological health, as well as richer and less easily quantified goods, would be desirable.

In addition to the wellbeing of patients, the impact of sociable robots on caregivers, designers, and patients' families and friends should be taken into consideration [18]. As noted earlier, using

robots in eldercare may help reduce the disproportionate burden of care work on some family members. But there are also costs. If, for example, using robots to satisfy elderly people's social needs causes us to devalue the elderly, this seems to count as a reason against their use even if the patients themselves do not feel the impact of this effect. If substituting robot interactions for human ones tends to weaken family ties, that could also affect people beyond the patient. If using robots to deceive patients into thinking they have friends when they do not makes us more comfortable with deception, we have reason to tread with caution in their design and implementation.

4. OF ECONOMIES AND FRIENDSHIPS

An analogy from Aristotle's discussion of deception and friendship provides a resource of the kind described above. But before introducing it, we should get clear on what exactly friendship consists in. The thought experiment involving Truman Show and Genuine options appealed to some intuitions on this front, but a more careful and clear articulation will prove useful in understanding what deception about friendship consists in.

Friendships can be taken to be objects in our social ontology – a kind of very small and close-knit group. Doing so offers a way to make sense of some common intuitions about friendship both within and without the Aristotelian tradition.

Friendship is often taken to involve unity and/or shared identity (consider, for instance Aristotle's oft-repeated claim that friends are "other selves"). But friendships seem as though they can be strengthened by differences, and accounts that interpret this shared identity as similarity face disadvantages [20]. Suppose one thinks of friendships and other close-knit social groups as objects in one's social ontology. If we construe friends as parts of composite objects – friendships – we can explain their shared identity in terms of parts of a whole rather than similarity between the friends. Inter-responsiveness and interdependence, as features of parts jointly composing a whole, then come to the forefront as characteristics of friends. This is consonant with many ordinary beliefs about characteristic qualities of friendship: that true friends reciprocate, that friends are characteristically those whose emotional states are responsive to their friends' wellbeing, and that friends' interests, broadly construed, include both their own wellbeing, that of their friends, and the wish that the two should remain interdependently connected.

Note that this does not preclude the possibility of similarities between friends, nor does it rule out similarity as one possible way that inter-responsiveness and interdependence can emerge and be sustained in friendship. It does, however, shift our focus, in thinking about friendship, from narrow consideration of the intrinsic features of individuals, to include features of the social groups they compose.

It has the explanatory advantage of accounting for the difference in intuitions between the Truman Show case and reciprocal caring, as the difference between real and merely-apparent social phenomena. In Truman Show, one is not appropriately interdependent with one's friends: the dependency runs only one way, as care and emotional responsiveness run in only one direction. In Genuine, by contrast, friends are mutually responsive to each other as parts composing a whole – there exists a genuine friendship, and not merely the appearance of one. Recall that according to Aristotle, friendships have intrinsic value. This interpretation supports the intuition that friends are not valuable merely for the experiences or other instrumental goods they provide. If friendships are social groups, then the groups themselves are valued, and not merely the external goods provided by affiliation.

I now apply this theory of friendship to unpacking an analogy Aristotle offers, in order to better understand concerns about the potential wrongness of using sociable robots to provide the subjective appearance of friendship without the existence of a grounding entity. In Book IX of the *Nicomachean Ethics*, following a discussion of friends as other selves and a detailed exploration of the reasons that conflict arises in various kinds of friendship, Aristotle comments on problems that arise between people when appearances of friendship fail, in various ways, to match reality. About this, he says the following:

> We might… accuse a friend if he really liked us for utility or pleasure, and pretended to like us for our character… if we mistakenly suppose we are loved for our character when our friend is doing nothing to suggest this, we must hold ourselves responsible. But if we are deceived by his pretense, we are justified in accusing him—even more justified than in accusing debasers of the currency, to the extent that his evildoing debases something more precious. [1]

It is worth noting that this account explains the badness of false friends not in terms of malicious intent, but representation of one thing as another. This is consistent with intuitions about the Truman Show scenario, where paid actors seem poor substitutes for friends even if they never betray one's trust or seek to use this illusion to harm the person they "befriend".

Rather, Aristotle's analogy between false friends and false coinage suggests that money constitutes membership in an economy. An economy, like a friendship, is a social group, albeit of a different kind. One can speak, for example, of the way the British economy reacts to a war, because an economy, like a friendship, is dependent upon and partly defined by the interdependence and inter-responsiveness of its members. Like friendships, economies derive their value from this. Unlike friendship, the value of an economy appears to be primarily instrumental, and one might opt for a world without economies if the same external goods were equally well realized by other means, unlike in friendship. But the analogy need not be perfect in order to be instructive. False money gives a false impression of membership in a social group, and its badness derives from this specific misrepresentation; likewise for friendship.

This suggests that there is something independently bad about counterfeiting, over and above the harm any given individual may or may not suffer in handling counterfeit currency. Consider, by way of illustration, the Case of the Compassionate Counterfeiter:

> *Compassionate Counterfeiter:* A is experiencing anxiety about money, which would be alleviated if A believed A possessed more of it. Out of concern for A, in order to assuage these worries, B writes A a bad check. As it happens, A never deposits the check, and eventually receives (actual) funds sufficient for financial support from a new job.

It seems to me that B has done something wrong in writing the bad check, even though A never cashes it. B's wrongdoing seems to consist in giving A the impression that A has connections to an economy that A could draw on, but which A does not in fact have. Economies, unlike friendships, are primarily instrumentally valuable, which may change the picture somewhat. Different goods may be implicated in the false appearance of each. But insofar as both friends and money constitute membership in a

social group which people find valuable, giving the false appearance of membership seems bad in similar ways.

This is not meant to imply either that patients are not also harmed, or have no reason to complain, when they are given the appearance of friendship without the reality. However, it suggests that the harm such patients suffer may be explained by the specific kind of deception: being given the impression that they are involved in something of value when there is nothing in which to be involved. We know, from the discussion so far, that we value friendships intrinsically. We have reason to believe that certain kinds of cognitive confusion prevalent in geriatric populations, in conjunction with our common vulnerability to certain features and gestures that elicit strong and often involuntary emotional reactions (such as big "eyes" and other anthropomorphic features), can cause the false appearance of friendship. We ought, then, to avoid exploiting these vulnerabilities in order to produce counterfeit friends.

5. COMPARING ARISTOTLE'S ANALOGY TO THE COMPETITION

Aristotle's analogy directs us to consider both whether people are fooled into thinking they are members of social groups, and who is to blame for the deception. This is helpful in identifying potentially relevant features of context in assessing the ethics of a given situation. The analogy introduces an account of the badness of sociable robots – not always an overriding badness, but a reason to be judicious in their design and use. It is helpful to consider how this account compares with other extant accounts of said badness, which can be thought of as attempts to explain the intuition that sociable robots can be ethically problematic. In what follows, I compare these accounts of badness in order to show why the version suggested by Aristotle's analogy is superior to others on offer.

Some accounts focus on the putative badness of sociable robots' capacity for enchantment. They hold that because sociable robots "enchant" by appealing to social instincts, they thereby deceive us, even against our better judgment. [17, 6] However, the proponent of an Aristotelian counterfeiting analogy can respond that this seems problematic. First, it makes seemingly self-aware testimony of seniors who report benefits of interaction with sociable robots to be unreliable, which runs the risk of paternalism. Some seniors are cognitively compromised, but not all are. Respect for agency, and for patients' own judgments about what constitutes their good, ought to be part of responsible geriatric care. We ought not to override their expressed preferences simply because they disagree with ours. Secondly and more generally, it is implausible to think that "enchantment" is sufficient for badness: we voluntarily watch "tearjerker" movies and read comedic novels, both of which work by appealing to our social responses and eliciting powerful emotional reactions, despite our intellectual judgments that the characters in the story are not real. Such "deception" seems wholly innocent. More broadly, playful simulacra of real phenomena can enrich our lives. Stories engage our social responses. Monopoly "money" can engage our economic reasoning faculties. Even when enchantment is harmful, it may sometimes be plausible that blame should fall on the user, as Aristotle cautions us, and does not necessarily show that anything is wrong with the object. The fact that some patients may misjudge their own social needs is not, in itself, reason to prohibit use of sociable robots altogether. Consider that adults of all ages sometimes make ill-advised choices when it comes to friends and companions; this does not license others to run their social lives. Enchantment thus seems an unsatisfactory explanation of the badness of sociable robots.

Another kind of account focuses on robots' potential to substitute for human relationships as the ultimate source of their badness. "Substitution" accounts say sociable robots are bad when and because they substitute for interaction with human beings [12, 14]. But the counterfeiter can respond that it is an empirical question whether robots will substitute for human interaction or facilitate it [18]. For instance, Wada and Shibata found that introducing Paro to nursing home residents increased both the quality and quantity of patients' interactions with each other, including a marked reduction in what they characterized as "backbiting" [19]. This is not so surprising if we consider that social abilities may, like muscles, get stronger with practice. It may also be that, by using robots to address some of the more stressful parts of eldercare, people will find it more enjoyable (and hence be more likely to make it a priority) to spend time with elderly friends and relatives. It is not wildly implausible to think that elderly relatives who are less depressed and more stimulated will be more pleasant and less draining to be around, thus creating a virtuous circle whereby they are included in more social events, and hence less likely to suffer from the ill effects of loneliness.

It is also worth noting that not every substitution of robot for human interaction is bad. After all, not all human interaction is desirable [14], and robots may supplant interactions with abusive or insensitive caregivers, or even just meddling busybodies, people whose presence in a patient's life would not constitute a good [18]. But the substitution account faces another disadvantage: even without causing a decrease in human interaction, it seems intuitively problematic to mislead patients, and the substitution account cannot explain why.

This last intuition motivates a "deceptiveness" account of the badness of sociable robots. According to a generic deceptiveness account, sociable robots may be bad when they deceive people, because they deceive people. Benign deceptions (defined as good intentions plus good consequences) are possible but rare exceptions to this principle [6]. Because the counterfeiting account is a kind of deceptiveness account, it may seem odd for me to object to such accounts. But the generic deceptiveness account differs from the counterfeiting account in its level of generality. Deception is often bad, but the times when it is least problematic are precisely the sorts of cases for which sociable robots are designed. For instance, deceptiveness may seem – in some cases – permissible for paternalistic reasons, as when cognitive confusion makes a person a poor judge of their own good. As Grodzinsky et al note, it may also be justified when done to facilitate ease of use, as when a Graphical User Interface (GUI) designer calls something a "file folder" even though no physical folders are involved. One might think that in some cases, transparency would impede the robot fulfilling its intended function, especially with a population not familiar with cutting-edge technology. It would not be practical to give seniors a crash course in robotics before using robots for basic care. Metaphors can help non-technical users to successfully interact with advanced technologies.

Because we need to distinguish permissible from impermissible deception, and because there are instrumental benefits to deploying sociable robots, as detailed earlier, more detail is required. Clarifying the content of the deception can help us better evaluate individual tradeoffs, and the counterfeit account provides such clarification. Given that sociable robots can provide patients with the subjective experience of friendship, we cannot yet

determine whether or not this constitutes a benign deception. Rather, we need to ask, is giving someone the subjective experience of friendship a good thing to do? This is what the counterfeit account addresses.

6. APPLYING THE ANALOGY

The counterfeit account characterizes the badness of sociable robot use in the following way. Friendships as complex interdependent social entities are objectively valuable, and as human beings who thrive in such groups it is important for us to recognize this value. Fooling others into believing they are parts of nonexistent social groups both fails to provide them with what they would consider good lives on their own terms, and shows a failure to treat the institution of friendship as valuable.

Application of this account yields results for both designers of sociable robots, and caregivers who use them to provide care for geriatric patients. Designers should minimize confusion – and be clear about target audiences, as this will vary depending on the cognitive capacity of the patient. Caregivers need to consider specific needs of particular patients to provide instrumental benefits without causing confusion analogous to counterfeit currency.

It also cautions us to be moderate and sensitive to context in our assessments of badness. Not every use of sociable robots constitutes an analogy to counterfeiting, and not every social good requires genuine friendship. We should recognize that the appearance of friendship can "toy with" our emotions without being morally bad (compare to the "enchantment" of films and stories). As discussed earlier, the subjective experience of friendship provides other goods, such as physical health and alleviation of perceived loneliness.

It is also relevant whether a given object is more plausibly taken to falsely imply a relationship where there is none, or to extend an existing relationship. Telephones, for example, "extend" real relationships by allowing people to interact despite lack of physical proximity, and email "extends" relationships by facilitating temporally discontinuous communication. Neither would count as counterfeiting. Similarly, other uses of technology might also constitution extension of real relationships rather than false implication of a non-existent relationship. For example: Sharkey and Sharkey relate a story of a woman who made an audio-recording of herself, reassuring her father and telling him to go back to bed. This recording was connected to a motion sensor in the man's home, and would play when he got up and began wandering about the house in the middle of the night [13]. This might plausibly seem more like an extension of their real relationship that helps her to achieve uninterrupted sleep while still reassuring her father, than a phony impression of a non-existent relationship. There may be cases where robots might be used to reassure confused patients of their real social connections.

One might think that this implies that *any* use of a sociable robot could be justified, provided someone somewhere along the line from design to production to deployment genuinely cared for seniors. The reasoning could be something like: Dee the Designer cares about seniors, so anything Dee designs will extend Dee's actual concern, so anyone who befriends a robot Dee designed is actually engaged in an extended friendship with Dee. But this will not serve as a blanket justification for using sociable robots in settings where caregivers' or designers' care is for patients considered generically, rather than particular patients with whom they engage in particular close relationships. The designers of a sociable robot could not say, for instance, that because they care for seniors, their design constitutes an extension of their friendship with any patients who happen to interact with the robot.

Friendships are highly particular relationships. They are not generic and not fungible. Friends cannot be substituted for one another: when one befriends a new person, one establishes a new friendship. So this account will not suffice to justify generic design features, even otherwise valuable ones. An example may help clarify. Money is valuable but highly fungible, and any particular instance of the generic kind "U.S. dollar bill", for instance, can (insofar as it is considered as money) be freely substituted for any other. But giving money as a gift is often considered problematic in personal relationships, precisely because of its generic and interchangeable nature. Contrast this with the significance of giving a friend or family member a handmade item: even if ugly or otherwise lacking in generic value, it often seems especially valuable because it expresses one's unique connection to a particular other person. So a generic concern for seniors will not make all sociable robots associated with such concern genuine friends.

7. CONCLUSION

Aristotle's analogy between counterfeit currency and false friendship thus provides guidance on the ethical use of sociable robots, without claiming to offer a uniform solution that glosses over important details of context. By characterizing the badness of sociable robots (when there is badness to be characterized) as giving the false impression of membership in a nonexistent social group, we can minimize the risks of said harm. There remains the possibility that the harms of counterfeiting friendship may be outweighed by other considerations, just as other kinds of paternalism might be justified in some circumstances. And as we saw earlier, some benefits associated with sociable robot use are quite powerful. But this seems an advantage rather than a drawback of the theory. Providing good care for the elderly, especially those whose cognitive and physical disabilities significantly impair their ability to live good lives, is a complex and demanding problem that will occasionally require tradeoffs between competing goods. Nonetheless, the analogy remains useful in helping us to correctly evaluate what goods are in competition, and when.

8. ACKNOWLEDGMENTS

Thanks to Randall Landau, Richard Volkman, Eric Ciardiello, Justin Grey, and an audience at San Jose State University for their feedback on early versions of this project, and Heidi Howkins Lockwood for her encouragement to develop it from a brief comment into a paper.

9. REFERENCES

[1] Aristotle. 1999. *Nicomachean Ethics*. Terence Irwin, Trans. Hackett, Indianapolis, IN.

[2] Banks, M. R., Willoughby, L. M., & Banks, W. A. 2008. Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *Journal of the American Medical Directors Association* 9 (March 2008), 173-177.

[3] Bynum, T. 2006 Flourishing ethics. *Ethics and Information Technology* 8 (Nov 2006), 157-173.

[4] Borenstein, J., & Pearson, Y. 2010. Robot caregivers: harbingers of expanded freedom for all? *Ethics and Information Technology*, *12* (September 2010), 277-288.

[5] Fratiglioni, Laura, et al. 2000. "Influence of social network on occurrence of dementia: a community-based longitudinal study." *The Lancet* 355.9212 (2000), 1315-1319.

[6] Grodzinsky, F. S., Miller, K. W., & Wolf, M. J. 2015. Developing automated deceptions and the impact on trust. *Philosophy & Technology* 28 (March 2015), 91-105.

[7] Kanamori, M., Suzuki, M., & Tanaka, M. 2002. Maintenance and improvement of quality of life among elderly patients using a pet-type robot. *Japanese Journal of Geriatrics* 39 (June 2002), 214–218.

[8] Misselhorn, C., Pompe, U., & Stapleton, M. 2013. Ethical considerations regarding the use of social robots in the fourth age. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry* 26 (June 20 13), 121.

[9] Parks, J. A. 2010. Lifting the burden of women's care work: should robots replace the "human touch"? *Hypatia* 25 (March 2010), 100-120.

[10] Ponte, E. 2014. Recent FDA medical device regulation and its relevance to robotics. *Tech Policy Lab* (January 27, 2014) http://techpolicylab.org/fda-medical-device-regulation/

[11] Sharkey, A. 2014. Robots and human dignity: a consideration of the effects of robot care on the dignity of older people. *Ethics and Information Technology* 16 (March 2014), 63-75.

[12] Sharkey, N., & Sharkey, A. 2010. Living with robots: Ethical tradeoffs in eldercare. In *Close engagements with artificial companions: Key psychological, social, ethical and design issues*. Yorick Wilks. Ed. John Benjamins Publishing Company: Amsterdam. 245-256.

[13] Sharkey, A., & Sharkey, N. 2012. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology* 14 (March 2014), 27-40.

[14] Sorell, T., & Draper, H. 2014. Robot carers, ethics, and older people. *Ethics and Information Technology* 16 (September 2014), 183-195.

[15] Sparrow, R., & Sparrow, L. 2006. In the hands of machines? The future of aged care. *Minds and Machines* 16 (May 2006), 141-161.

[16] Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, A., Higashi, Y., et al. (2004). Is an entertainment robot useful in the care of elderly people with severe dementia*? Journals of Gerontology Series A Biological Sciences and Medical Sciences* 59 (January 2004), 83–85.

[17] Turkle, S. 2011. *Alone Together*. Basic Books, New York, NY.

[18] Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. *Philosophy & Technology* 24 (September 2011), 251-268.

[19] Wada, K., & Shibata, T. (2006) Robot therapy in a care house: Its sociopsychological and physiological effects on the residents. In *Ethical issues in robot care for the elderly* 39 123 Proceedings of the 2006 International Conference on Robotics and Automation, Orlando, Florida (May 2006), 3966–3971.

[20] Williams, B. 1981. Persons, character, and morality. In *Moral Luck.* Cambridge University Press, Cambridge, UK.

[21] Wilson, R. S., et al. 2007. Loneliness and risk of Alzheimer disease. *Archives of General Psychiatry* 64 (February 2007), 234-240.

Trusting the (Ro)botic Other: By Assumption?

Paul B. de Laat
University of Groningen,
Groningen, the Netherlands
p.b.de.laat@cerug.nl

ABSTRACT

How may human agents come to trust (sophisticated) artificial agents? At present, since the trust involved is non-normative, this would seem to be a slow process, depending on the outcomes of the transactions. Some more options may soon become available though. As debated in the literature, humans may meet (ro)bots as they are embedded in an institution. If they happen to trust the institution, they will also trust them to have tried out and tested the machines in their back corridors; as a consequence, they approach the robots involved as being trustworthy ("zones of trust"). Properly speaking, users rely on the overall accountability of the institution. Besides this option we explore some novel ways for trust development: trust becomes normatively laden and thereby the mechanism of exclusive reliance on the normative force of trust (*as-if* trust) may come into play - the efficacy of which has already been proven for persons meeting face-to-face or over the Internet (virtual trust). For one thing, machines may evolve into moral machines, or machines skilled in the art of deception. While both developments might seem to facilitate proper trust and turn as-if trust into a feasible option, they are hardly to be taken seriously (while being science-fiction, immoral, or both). For another, the new trend in robotics is towards coactivity between human and machine operators in a team (away from making robots as autonomous as possible). Inside the team trust is a necessity for smooth operations. In support of this, humans in particular need to be able to develop and maintain accurate mental models of their machine counterparts. Nevertheless, the trust involved is bound to remain non-normative. It is argued, though, that excellent opportunities exist to build relations of trust toward *outside* users who are pondering their reliance on the coactive team. The task of managing this trust has to be allotted to human operators of the team, who operate as linking pin between the outside world and the team. Since the robotic team has now been turned into an anthropomorphic team, users may well develop normative trust towards them; correspondingly, trusting the team in as-if fashion becomes feasible.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Ethics

General Terms

Human Factors

Keywords

Artificial agents, coactivity, institutions, men-machine team, mental modelling, trust

My plane has just landed at the airport, late at night. Outside a taxi is waiting for customers. Unfortunately, all I have are 100 dollar notes. Should I trust the robotic driver *to take me to the distant location that I have to go to? More precariously, should I trust this robot to change one of my notes into local currency and give me back the change?*

1. INTRODUCTION

Three decades ago trust was a concept that only pertained to relations between persons. Since then the world has seen a series of technological developments in ICT that have affected and transformed social life profoundly. As a corollary it has been found necessary to enlarge the domain in which the conception of trust can be applied. For one thing, we have to consider e-trust or virtual trust, concerning the relations between people over the Internet. For another, we have to consider relations between human agents and artificial agents (be they robots or bots) – and even between such artificial agents themselves. Also in the case of such relations we have to pose the question of trust.

Obviously, enlarging the domain to which trust applies is not without its problems. The mediation by the Internet as well the silicon-embodied nature of artificial agents complicates the issue of trust considerably. Is it justifiable to continue talking about trust?! Sidestepping these questions for the moment, let us follow the example of Grodzinsky *et alii* [1] who introduced the overarching conception of TRUST, which as a superclass contains attributes common to all domains just discussed. Although, as the authors say, ''ethically significant differences'' exist between the various applications, let us subsume them under the larger umbrella of TRUST.

In order to arrive at this umbrella concept, the authors had to rephrase the usual conception of trust. As a rule, face-to-face trust and virtual trust are defined as reliance on the good intentions of the trustee in situations of dependence, vulnerability, and high risk for the trustor. Now, with artificial agents playing their roles, these good intentions had to be abstracted away. Instead, they introduced the 'expectation of gain' – the trustor expects to gain something valuable by trusting the trustee. As long as a trustee delivers, it does not matter whether good intentions, self-interest, lines of code - or whatever - are behind this fulfilment; the expectations involved in TRUST are *no longer normative*.

2. TRUSTING *AS-IF*

The mechanisms bringing people to trust other people (face-to-face trust) have always been varied. We may infer other persons' trustworthiness from their individual characteristics, their self-interests in the situation, their belonging to a shared culture, or their membership of a trusted institution. Perceived reputation is also a powerful motivator. If interaction is more continuous, trust may be produced by the introduction of monitoring mechanisms in the background of the community, or rules of hierarchy in the foreground (for all this cf. [2]). One more mechanism though is always a possibility as well: trust is not inferred in any way, but *assumed* at the outset. The precondition for any trust to evolve (trustworthiness) is a matter of assumption, not of inference. We act *as if* trustworthiness is present; by acting as if we (hopefully) *produce* the trust that we seek. The mechanism has first been signalled by authors like Gambetta [3], and Luhmann [4].

The force of this mechanism is not to be underestimated: it may lead to trust in new and unknown situations where all usual indicators for trustworthiness are lacking. Instead of waiting for tangible proof of trust, usually in a series of reciprocated gestures of trust, the potential trustor jumps to the conclusion that trust is warranted (in spite of the lack of evidence). As a result, the development of trusting relations that otherwise would need much more time gets jumpstarted and accelerated – on condition, that is, that the gesture is reciprocated indeed.

This face-to-face mechanism is puzzling to say the least. Why would anyone rely on it? With any act of trust, a *normative* force comes into being: it imposes a normative claim on the trustee to respond in like fashion (while being *eine supererogatorische Leistung;* cf. [4]: 55). With the mechanism of assuming trust, the trustor chooses to rely on this normative force *exclusively* – lacking any other indicators for trust. Two strands of theory have attempted to give the mechanism more rational substance. The first is from Pettit [5] who hypothesized that I may believe that the other cherishes my esteem and will therefore reply in kind to a gesture of trust (so called trust-responsiveness). This gamble will pay off as long as the other does not want to forfeit the chance to reap my esteem. So the central element here is the seeking and giving of esteem.

Later on, in a critique of the ''cynical'' and calculating character of the esteem mechanism as proposed by Pettit, McGeer proposed the mechanism of substantial hope in the capabilities of the other [6]. The trusting gesture is an appeal on the other to develop their capabilities to the full. They are offered the chance to empower themselves. So in contrast to Pettit where other people are seduced in cunning fashion, people are openly and without subterfuge challenged to prove what they are worth.[1]

As-if trust is undoubtedly a mechanism to be reckoned with in face-to-face trust. Think of the new resident that gives his key to the neighbour while he takes a vacation, or the passenger that trusts her taxi driver to take the shortest route possible ([5]: 218). Since then it has been argued to be a sizeable force in situations of virtual trust as well.[2] Also when we meet unknown others over the Internet, we may be inclined to give them the benefit of trust. Examples that have been mentioned in this regard are many (cf. [7] [2]). Members of self-help groups and online diarists expose intimate details of their lives. Open content communities do likewise. When collaborating on open source software, they open up their source code repository for modification by others. When working together to produce an encyclopedia (like Wikipedia), they entrust their entries to outside contributors at large. Although the virtual others are invisible and unknown, and could easily abuse what is entrusted to them, investing trust in them is chosen as the default. As a result, trust may flourish very quickly and benefit the collaboration immensely.

3. TRUSTING ARTIFICIAL AGENTS *AS-IF*

We now turn to relations of trust (TRUST) between humans and artificial agents. Can the mechanism of as-if trust play the same catalytic role in them as in face-to-face and virtual trust relations? Immediately we encounter the problem that artificial agents are not the most likely candidates for such trusting-by-default. Let us focus on state-of-the-art artificial agents, like the ones discussed in Tavani and Buechner (forthcoming): artificial agents that are rational, interactive, adaptive, and independent, together designated as ''functionally autonomous''. An example in point is the robotic Johnny Bot that manages to navigate your car through all traffic in autonomous fashion [8]. It may make sense to talk about trusting this robot before getting into one's car. In analogy to trusting a human driver, we may put some trust in the robotic driver. This trust, however, can only be of the TRUST category in general: we expect the robot to deliver us safely at our destination – but these expectations are *not* of a normative kind. We do not hold the bot responsible for safe arrival (as we do any human driver).[3] [4]

If this interpretation is accepted, relations of trust between human and (sophisticated) artificial agents have a hard time to develop. A normative appeal on the (ro)bots is destined to fall on deaf ears.[5] So at first sight it would seem that only one dynamics is possible: human agents test the ''intentions" of their artificial counterparts in small steps. To the extent that trust is answered with trust, ever larger steps of trust can be taken.

1 Note that Annette Baier, the 'arch-mother' of the revival of trust research in the 1980s, never got round to accepting the validity - let alone the force - of as-if trust. She argued that, for one thing, hope is not a proper support for trust, only vice versa. For another, as-if trust is a form of moral pressure that manipulates or (at best) seduces – and as such cannot be accepted, but had rather be resisted [20].

2 Although Pettit completely *denies* the possibility of virtual trust – let alone of any rational assumption of virtual trust [21].

3 So here I disagree with Tavani and Buechner who operate with a normatively laden conception of trust and assert nevertheless that such trust *is* possible between human agents and sophisticated artificial agents [10].

4 Some critics will object that the designer(s) in the background have to be taken into account. After all, we may harbour normative expectations towards them, so trust them in the usual sense of the word. But do we humans really take this chain of causation into account when we gauge the reality of our relations to the machines we meet? To complicate matters, appealing to the designers' good intentions in a move of as–if trust might only bring design improvements later on, not instantly a trusting response from the robot that they designed.

5 Note though that Hoffman *et alii* suggest the possibility that novices with technology might be naïve and take to swift trust – to be disappointed soon [18].

In spite of this I would argue that realistic future possibilities do exist for as-if trust to accelerate the dynamics of developing trust between human and artificial agents. Before exploring them, however, it has to be emphasized that humans do not necessarily meet robots in isolated encounters. Instead, the robots involved may be embedded in an institutional context of men and machines – so-called *"zones of trust"* (cf. [9]; [10]; conception adopted from Walker [11]). Think of supermarkets being stocked by an array of robots, banks relying on bot procedures over the Internet, or airline companies using an array of automated machines in their planes (up to robotic pilots). Even if we do not meet the robots involved personally, we may ex ante have trust in the institution as a whole that they perform as we desire from them. If such institutional trust obtains, we have a motive to engage in particular acts of trust towards them (say to go shopping, pay over the Internet, or take our seat in the plane). The trust is not invested in particular persons or particular robots – we are just confident that somewhere somehow the institution can be relied upon ([11]: 85).

Let me in passing make this interpretation (as proposed by Walker) more precise and carry it to its logical conclusion: we are confident that the institution can be held to *account*. That is, in an interesting twist, such trust is no longer tied to individual responsibility but to collective accountability (cf. [12]). Thus the problem of which persons or robots are to be blamed (the "many hands" involved) is circumvented. If institutional trust exists, the robots on board take a free ride so to speak on our abilities to trust. Since we trust the institution, we also trust the robots they happen to have incorporated in their ''workforce''. Trust transfers from the institutional zone as a whole to the collective of actors within, be they carbon or silicon based.

This theorizing about zones of trust of course begs the question as to how mutual trust between human and artificial agents has been established *within* the institution in the first place. How did trust "originally" evolve between them? Was it only the slow process as associated with a gradual increase of non-normative trust (TRUST)? Or did the mechanism of pure assumption of trust accelerate the process? On closer inspection some possibilities for as-if trust to be relied on do seem to be available, at least in the future.

One possibility is that robots develop *morality by design* [13]. Moral capabilities that suit the problems at hand are built into the machine. Their morality is just functional, not moral. As to their emotions, they will have no ''real" but synthetic ones. Robots develop into "quasi-others" for us, simulating moral agency ever better [14]. If this scenario realizes, humans may have some faith in the mechanism of as-if trust since the machines involved may answer our emotional appeal in synthetic fashion. But at present, this sounds more like science-fiction than reality.

Another, more realistic possibility of the kind is, that (ro)bots become masters of *deception* [15]. Designers learn how to build deception into the machine. They masquerade as humans and manipulate us to consider them as made of flesh and bones. As a result, humans may be tempted to trust them without any indication that such trust is warranted: they just assume the (deceiving) robots can be trusted. This jumpstart mechanism, though, is both immoral and unreliable. From a moral point of view, as Grodzinsky *et alii* [15] convincingly argue, the default must be that robots should not be programmed to deceive. And from a practical point of view, in the long run the deception may be discovered and trust unravels as a result.

There seems to be, however, another more realistic and more reliable possibility for trust to evolve between human and artificial agents. It consists of welding the robots involved together with humans in a *team* that collectively addresses the tasks at hand. Humans graduate from operators in the background to fully involved operators. Thereby prospects are opening up, after all, for humans-as-users to rely on direct as-if trust in their dealings with the machine-in-a-team. The details of this trend towards the formation of teams are explained in the sequel.

4. COACTIVE TEAMS

Traditionally, robotics has been geared towards the end-goal of eliminating all human support. Fully automated factories and hospitals were the utopic vistas. The messy ''social dimension'' had to be replaced by reliable machine procedures. Gradually one has come to realize that this goal is an illusion. One of the main reasons is, that contingencies are part and parcel of most sophisticated activities; these may easily upset a fully automated design. Therefore some designers have to come to advocate a new approach. This new trend in robotics is to build teams for performing the tasks at hand that consist of both robots and men in close cooperation; the core conception being *coactivity* [16]. Human and artificial agents each have their specific capabilities; these are combined in such a way that the best possible performance is the outcome. Crucially, this is not robots doing the job, with humans stepping in where they fail. On the contrary, both kinds of agents fulfil complementary roles. The strong points of each of them are being exploited.

Although less sophisticated tasks (like mowing the lawn) may remain the province of autonomous robots on their own, many designers in the field believe that sophisticated tasks to be accomplished by means of robots are better carried out by such mixed teams. Examples mentioned are as diverse as robots for assisting the elderly and/or disabled, hospital robots, and search-and-rescue teams in war zones or remote areas. The authors of [16] provide an example of such a robotic team as they designed it themselves. The DARPA Robotics Challenge has been created to spur the development of advanced robots that can assist in the recovery of humans from a disaster. As part of this challenge a software competition took place in a virtual environment that looked like a suburban area. A simulated robot, operated from the distance, had to complete a range of tasks (like walking over a hill, driving, picking up a hose). The human operator from the distance performed crucial sensing tasks by using a data interface. After 56 hours of simulation the robotic team from the authors took first place.

At the heart of coactive design stands the task of designing for the interdependencies involved ([16]: 49). The most obvious ones are "hard", having to do with the more material aspects of the joint tasks. But there are also "soft" interdependencies, relating to helpful observations and actions that are supportive of the process (but not strictly required). Think of observations about how an actor is doing, about unexpected dangers that are observed ahead, and helpful suggestions for support. These important soft interdependencies can be subsumed under the more general category of "attention management" (cf. [17]: challenge 9): 'Team members [must] direct each other's attention to the most important signals, activities, and changes.'

Interestingly, workers in the field of robotics have become aware that for men-machine systems ''new style'' to be resilient, mutual trust between them is indispensable [18]. If trust reigns in men-

machine systems, an – often tacit - "basic compact" between men and machines can be said to be in force, in which partners agree to work towards mutual goals ([17]: challenge 1). Essential elements of such mutual cooperation are responsiveness and reciprocity.

As an essential precondition for mutual trust, the factors governing the machines and the state they are in should at all times be visible and available for inspection to the human operators involved [18]. Since these operators, as humans, are bound to make "mental models" of the machines involved, these models had better be accurate in order to smooth relationships. It is on this point that coactive teams have much more promise than (semi-)-autonomous machine systems. In the latter, the designers' intentions are usually black-boxed and inaccessible; in the former, they are – or at least can be - accessible which as a rule facilitates ''mental modelling" and thus promotes mutual trust.

In order to explicitly facilitate and nurture men-machine trust, researchers are now working on the design of techniques for "active exploration for trusting" (AET) [19]. These should enable monitoring the relations of trust between operators and their machines and keep them ''calibrated''. Such design has to face several challenges; let me just mention two of them. One hard problem is to conceive what kind of messages/indicators about trust the machines can deliver (so-called trust and mistrust 'signatures'). Another is that an operator should be enabled to actively probe the machine to test whether specific hypotheses about trust are warranted or not. The team is never only in operational mode, but always also in experimental mode.

5. MULTIPLE TRUST RELATIONS

With this new style robotic team, unavoidably the issue of trust has multiplied. We started with the issue of trust between human *users* and their machines, machines that were tacitly understood to be semi-autonomous. Now it transpires that in future human users may well meet their robots embedded in a team; in them, humans and machines cooperate as *operators*. And as we have just seen, also for the team relations between men and machine the issue of trust lies square on the table. We have ended up in a more complicated situation.

I would argue that, nevertheless, the prospects for trust to emerge and grow have become better. In particular the possibilities for as-if trust to jumpstart mutual trust emerging have increased. Let me explain. Consider first the issue of trust *within* the man-machine-team. As elaborated above, trust on the part of humans (of the more general category of TRUST) develops to the extent that they succeed in making an adequate mental model of their robotic counterparts and are enabled to stay informed throughout of where the robots ''stand" in the operational process. This development of trust is bound to be a slow process. Moreover, since the trust involved is non-normative (TRUST), as-if trust playing an invigorating role appears to be out of the question.[6]

Let us look broader now, towards the other issue of trust where humans as users or beneficiaries meet a coactive robotic team. It concerns sophisticated applications with users in need of help that (have to) rely on them. Think of patients in the hospital, whether physically or mentally ill, casualties in a battle zone, victims of an accident in the mountains or the jungle, and the like. I would argue that avenues are open for the users concerned to quite naturally approach such a team as an animated entity, not as a machine entity. Such avenues are open, if the opportunity for such a meeting of minds is explicitly taken up by the team and their designers. This necessary *enlargement* of trust management practices – which, to my knowledge, has up to the present been neglected by researchers in the field of robotics - should take shape in the following manner.

As a supplement to trust management of the relations inside the men-machine-team, the management of trust relations towards the outside users/beneficiaries should be explicitly undertaken. Designers have to realize that users most usually will be hesitant to trust. Everything is to be done to assuage their fears. For the purpose it is immensely helpful that humans are present as members of the team. It is to some of them – and *not* to some of the machines – that trust management tasks are to be allotted. Moreover, these "trust officers" should not just be public relations officers, but fully functioning members taking part in the operations themselves.

With these efforts in place to manage trust towards outside users, these may easily make a mental model of the team since it just amounts to modelling the human operators as the gateways to the team – and such models we are used to making on a daily basis, whenever we meet other people.[7] Therefore, since the robot has been transformed into a more anthropomorphic team, the relations of TRUST have transformed into relations of trust proper, allowing expectations of a normative kind.

Correspondingly, ordinary humans may more readily take to trusting the team without any indicators being available, thereby purely relying on the normative force of trust. Trust is assumed, not inferred in any way. We may take the chance and rely on the normative force of trust plain and simple, while the keepers at the gate look quite like us. They will understand our appeal alright.

Observe that human team members may have a close or a more distant location relative to the users involved, depending on circumstances. Consider the application that a robot is missioned to rescue victims of an accident. If they have to be rescued from a traffic accident in a city, it seems most sensible to include the human operators in the team that actually hurries to the location of the accident. If the victims, however, have fallen into a gorge in the jungle, safety would dictate to leave the operator somewhere outside the accident zone and let him cooperate from the distance. I would argue that the latter case presents a dilemma for trust. For reasons of safety, the human operator is preferably kept away from the physical place of the accident. The establishment of trust thus becomes more complicated: victims have to trust the human team member that they can only communicate with via some interface. In conclusion: physical proximity of the human operator of a coactive team would seem to be a requirement for the development of as-if trust in unencumbered fashion.

6 Let me mention as an aside that Hoffman *et alii* ask the question whether a machine can inspire 'swift' trust in it on the part of the human operators ([19]: 87). For the authors it remains a challenge to be solved. Note though, that their term 'swift' trust does not exactly square with the 'as-if' trust that is central to this article.

7 Note that this modelling is actually much easier than in the former case of the coactive team itself - of human operators modelling their *machine* counterparts.

6. CONCLUSIONS

This article has been an exercise to inquire into the conditions under which the resourceful mechanism of as-if trust could play a role to accelerate the development of mutual trust in this cyber- and robotic era. Already being operative in real life (face-to-face trust) and in virtual life (e-trust), can anything be expected from the mechanism when human agents meet artificial agents of considerable sophistication? At first sight the answer seems to be no. Trust towards artificial agents can only be non-normative, since they harbour no intentions but just embody those of their designers. As a corollary, appealing to such intentions and keeping them responsible is not imaginable as a rational option; trusting as-if would be plainly irrational.

Nevertheless it has to be kept in mind that many an artificial agent is first tried out in the back corridors of institutions small or large. After beta-testing with the help of advanced users wherever possible, they are tested by the members/users of the institution. Only when they are developed to the point that the institution can vouch for their functioning, they ate 'let loose' on the public at large. In that case, users confront artificial agents in a so-called 'zone of trust'. If they happen to trust the institution, they are going to assume that all actors inside contribute to an overall trustworthy performance – (ro)bots included. Notice that the normative claim of trust no longer attaches to individuals but to the institution as a whole; accountability is the central concept. Although this is - obviously - not an issue of humans resorting to blank trust, their relying on institutional trust is an important option towards developing full-blown trust towards artificial agents in a fashion.

This theorizing about zones of trust of leaves the question unanswered as to how mutual trust between human and artificial agents may come to be established *within* the institution in the first place. Can only the gradual increase of mutual TRUST be involved, or (as-if) trust as well? Some positive answers for the future that create room for as-if trust have been explored. First, machines may evolve into moral machines, by design. For now, this seems to belong to the realm of science-fiction. Secondly, robots may be designed to practice deception. In particular, they may become able to lure humans into believing that they are made of flesh and bones. Though more feasible, this option would seem to backfire in the end; moreover, to be morally unacceptable.

Finally, the new trend in robotics is away from autonomous machines. Instead, humans and robots are welded together in a team in which the capabilities of both kinds of operators are utilized in optimizing fashion (coactivity). Such teams need mutual trust between all actors. As a precondition for trust developing, human operators need to be able to maintain mental models of their machine counterparts, fed by accurate and current data about the states the machines are in. The development of trust within this coactive team can only be slow; any mental model of robots does not and cannot include intentions and emotions. Trust can only be non-normative.

This trend towards coactivity, however, has the potential to profoundly transform the relations between the team and the human users approaching it from the outside. If, and only if, management of the trust towards users of the collaborative work-system is put on the agenda and allotted to some of its human operators, users may more easily approach the system as anthropomorphic. The process of mentally modelling the ensemble involved transforms potentially from mere (non-normative) TRUST to (normative) trust. If this transformation takes place, the option for human users to take a chance and just assume that trust obtains becomes more realistic. Building of mutual trust may thus be accelerated. A requirement seems to be that the human operator functioning as linking pin for trust is situated as close as possible to the actual operations.

My plane has just landed at the airport, late at night. Outside a taxi is waiting for customers. The car is driven by a team*: a robot behind the wheel and a human operator behind a control panel. Unfortunately, all I have are 100 dollar notes. Should I trust the team to take me to the distant location that I have to go to? More precariously, should I trust this team to change one of my notes into local currency and give me back the change?*

7. REFERENCES

[1] F. Grodzinsky, K. Miller and M. Wolf, "Developing artificial agents worthy of trust: "Would you buy a used car from this artificial agent?"," *Ethics and Information Technology,* vol. 13, no. 1, pp. 17-27, 2011.

[2] P. B. de Laat, "How can contributors to open-source communities be trusted? On the assumption, inference, and substitution of trust," *Ethics and Information Technology,* vol. 12, no. 4, pp. 327-341, 2010.

[3] D. Gambetta, "Can we trust trust?," in *Trust: Making and breaking cooperative relations*, Oxford, Blackwell, 1988, pp. 213-237.

[4] N. Luhmann, Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität, 4 ed., Stuttgart: Lucius & Lucius, 2000.

[5] P. Pettit, "The Cunning of Trust," *Philosophy and Public Affairs,* vol. 24, no. 3, pp. 202-225, 1995.

[6] V. McGeer, "Trust, hope, and empowerment," *Australasian Journal of Philosophy,* vol. 86, no. 2, pp. 237-254, 2008.

[7] P. B. de Laat, "Online diaries: Reflections on trust, privacy, and exhibitionism," *Ethics and Information Technology,* vol. 10, pp. 57-69, 2008.

[8] H. T. Tavani, "Levels of Trust in the Context of Machine Ethics," *Philosophy and Technology,* online first.

[9] J. Buechner and H. T. Tavani, "Trust and multi-agent systems: applying the ''diffuse, default model'' of trust to experiments involving artificial agents," *Ethics and Information Technology,* vol. 13, no. 1, pp. 39-51, 2011.

[10] H. T. Tavani and J. Buechner, "Autonomy and Trust in the Context of Artificial Agents," Berlin, forthcoming.

[11] M. U. Walker, Moral Repair: Reconstructing Moral Relations after Wrongdoing, Cambridge: Cambridge University Press, 2006.

[12] H. Nissenbaum, "Accountability in a computerized society," *Sceince and Engineering Ethics,* vol. 2, pp. 25-42, 1996.

[13] G. D. Crnkovic and B. Cürüklü, "Robots: ethical by design," *Ethics and Information Technology,* vol. 14, no. 1, pp. 61-71,

2012.

[14] M. Coeckelbergh, "Moral appearances: emotions, robots, and human morality," *Ethics and Information Technology,* vol. 12, no. 3, pp. 235-241, 2010.

[15] F. S. Grodzinsky, K. W. Miller and M. J. Wolf, "Developing Automated Deceptions and the Impact on Trust," *Philosophy and Technology,* online first.

[16] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. van Riemsdijk and M. Sierhuis, "Coactive Design: Designing Support for Interdependence in Joint Activity," *Journal of Human-Robot Interaction,* vol. 3, no. 1, pp. 43-69, 2014.

[17] G. Klein, D. D. Woods, J. M. Bradshaw, R. M. Hoffman and P. J. Feltovich, "Ten challenges for making automation a 'team player' in joint human-agent activity," *IEEE Intelligent Systems,* vol. 19, no. 6, pp. 91-95, November/December 2004.

[18] R. R. Hoffman, J. D. Lee, D. D. Woods, N. Shadbolt, J. Miller and J. M. Bradshaw, "The dynamics of trust in cyberdomains," *IEEE Intelligent Systems,* pp. 5-11, November/December 2009.

[19] R. R. Hoffman, M. Johnson, J. M. Bradshaw and A. Underbrink, "Trust in Automation," *IEEE Intelligent Systems,* vol. 28, no. 1, pp. 84-88, January/February 2013.

[20] A. C. Baier, "Putting Hope in its Place," in *Reflections on How We Live*, Oxford, Oxford University Press, 2010, pp. 216-229.

[21] P. Pettit, "Trust, Reliance, and the Internet," *Analyse & Kritik,* vol. 26, pp. 108-121, 2004.

Robots make Ethics honest – and Vice Versa

Wilhelm E. J. Klein
School of Creative Media, City University of Hong Kong
18 Tat Hong Avenue
Kowloon Tong, Hong Kong
mail@wilhelmklein.net

ABSTRACT
This paper revisits major revolutions in human self-perception, and pursues their insights to their logical conclusions, using robots as conceptual archetypes for fully naturalistic, talking, walking and thinking agents. Doing so, humans are reconsidered as bio-bots and ontologically not of significant difference from techno-bots; morality is stripped of metaphysical remnants of the past and updated to a preference-utilitarian morality[2], and moral agency re-examined in light of a determinism and the non-existence of free will. Taken together, this robot-catalysed level of philosophical honesty provides a sound foundation for the task of making robots ethical.

Categories and Subject Descriptors
K.4.m [**Robot Ethics**]: Miscellaneous

General Terms
Theory

Keywords
robot ethics, naturalism, moral intuition, free will, determinism, moral agents/patients

1. INTRODUCTION
In the spirit of Dan Dennett's famous 'AI makes philosophy honest' [18], this paper argues that robots can make ethics honest, and philosophical honesty can make them ethical[1]. Dennett, of course, referred to the possibility of artificial intelligence necessitating honesty about our own consciousness, that we too are computing machines (albeit very complex ones) and no better or worthier or more special in an ontological sense than any other computing machine. Accordingly, philosophy of mind, he argued would have to abandon dishonest notions of brain-mind dualism and any other super-natural claims not grounded in naturalism, or as he coined elsewhere, "no sky hooks allowed" [16].

I believe a similar case can be made for (robot) ethics, as robots necessitate moral philosophers to be honest about the nature and origins of ourselves and our concepts and behaviour. We need to let go of notions of divine law, intrinsicality, metaphysical moral truths, freedom of will any many other such 'sky hooks'. Robots, as fully naturalistic and deterministic automatons represent the archetype of naturalistic, walking & talking agents without any added essences, elan vital, soul or anything else, thus function as an intuition pump to overcome our own cognitive bias and perception of specialness. It allows us to arrive at a "honest ethics". This, in turn, provides a conceptual basis that enables robot ethicists to bypass attempts to infuse robots with whatever it is that makes us special and moral as there is no need to get robots to the point where they are "morally responsible" or be "moral agents" who act out of free will etc. [3, 48, 35, 22], [30, ch.1].

2. HUMAN SELF-UNDERSTANDING
Human self-understanding has substantially changed through the centuries, and, it has been put forward, can roughly be categorised in certain, particularly transformative revolutions of thought [23]. Although quite west-centric, and as Freud himself acknowledged, reassessment of human nature is actually a very gradual process, they do serve as rough markers for certain ideas that could (and did) radically change our self-perception [23, ch.4].

As an introductory example, consider the Copernican revolution, as he delivered quite definitive proof that the earth is not, in fact, at the centre of the universe. This came as a shock to the people of his time, who, although they no longer believed in gods riding golden chariots across the sky, indeed still believed the earth to be at the centre of the universe (being put there by god). The foundation for such belief was usually not mathematics and astronomy as employed by Copernicus, but philosophical deduction based mostly on the mix of ancient Greek philosophy and holy scripture [50, **?**]. At the time, he faced quite a lot of criticism, who found his heliocentric model preposterous–not least because it took away one thing that made us special. But in time, and thanks to some other natural philosophers, we came to accept our place in the solar system, and a bit later even in our galaxy and universe.

[1]For related suggestions see [44, 4].

Maybe because, at least down on earth, we were still undeniably special. In fact, to any rational individual of the past view centuries, it was absolutely clear that humans are by far the most advanced being on the planet, clearly different from the rest of the animals, and, because anything complex obviously needs a designer, most certainly gods chosen people. It was only when Charles Darwin and his theory of natural selection came along that we had to reconsider these assumptions. We were forced to do so, because Darwin's (dangerous) idea enabled an perspective where all the richness we see in the natural world (including us) could have developed (evolved) completely without the need for a creator, divine designer, or anything else outside the natural world. Which means we too could be identified as an animal. Not "formed from dust" and divine, but a mostly hairless great ape with particularly evolved cognition. At his time, this kind of proposition was absolutely outrageous. Darwin himself hesitated for decades to publish his work because he feared the reaction of his contemporaries (and he was indeed right about the public outcry [20]). And in fact, not like generally accepted heliocentric model, many still find it difficult to overcome our *feeling* special, and continue to think and behave as if it was "humans and animals" not "humans and *other* animals".

But even those who accepted our place within the animal kingdom could at least rest knowing we were *the conscious animal*. The one that talks and thinks rationally. Until Sigmund Freud came along. So far we had thought we were the only ones who really own their mental contents, who have full access to their mental lives and are completely in charge of their own thoughts. Rene Descartes famously summarised this with his "I think therefore I am", mirroring the conviction of his contemporaries, who also considered the mind fully conscious and acting on the basis of reason. Freud's work in psychoanalysis and its focus on the unconscious shattered this belief as it revealed how much of our actions are influenced by our subconsciousness. Not only, it turned out, are we not aware of many things going on in our minds, these things also play a major role in forming our thoughts, ideas, dreams and actions. Today, this is generally accepted, albeit on a rather superficial level. We readily acknowledge how, say, a troubled childhood may lead to depression, but in everyday life, we continue to treat each other as if we were completely conscious, rational agents, and neglect the broader implications of the Freudian insight.

3. REVOLUTIONS CONTINUED

To a certain degree, this, along with the work-in-progress status of evolution acceptance may simply have to do with the passage of time. I still remember how silly it sounded to me that someone could have really doubted the heliocentric model when I was first taught about Copernicus. It seems that it has been burned into our common knowledge so deeply, that it simply *feels* silly to think otherwise. Not so much for the other two revolutions. When discussed, opposition to their propositions seem not nearly as silly as the one Copernicus faced. In 2014, Gallup reported that only 19% of US citizens believe in evolution with god having no part in it. The overwhelming majority still believes that god either created humans in their present form (42%) or at least had a significant influence in their evolution (31%) [37]. Similarly, on the surface, we appear to accept that we possess a subconsciousnesses, but very little does it seem to have really changed the way we see our selves and each other. We continue to fall victim to the fundamental attribution error and its cousins and judge ourselves by our intentions and others by their actions.

Within the ivory tower the situation appears better, but far from unanimous.In 2013, David Bourget and David Chalmers polled 931 philosophers and mapped their opinion on various topics [8]. They did not ask about the heliocentric model (why not?), but they did ask about belief in the external world (a whooping 81.6% believes there is one), metaphilosophy (49.8% naturalism, 25.9% non-naturalism and 24.3% other) and meta-ethics (56% moral realism, 27.7% moral anti-realism, and 15.9% others) suggesting that even in these circles, not everyone has taken the Darwinian and Freudian revolutions to their logical conclusions and embraced them fully.

This is the societal status quo, the Zeitgeist of the time in which some of us are attempting to make robots ethical. Accordingly, for some of those taking part in this endeavour, it is still difficult to accept a fully naturalistic, deterministic world, and thus they engage in attempts to conceptually infuse robots with something that appears to be only working with us special cases. How does one explain moral truth to a robot? And how could a fully deterministic machine-brain act out of free will, the very basis for being a moral agent? In the following, I will dig deeper into the Darwinian and Freudian revolutions, and outline why their extended implications and conclusions are relevant to such questions and the quest to make robots ethical.

3.1 Honest Origins

The not quite century-and-a-half since the publication of *The Descent of Man* has seen a lot of further scientific progress, including having learned about the particles that make up everything and the structures they can form to become the molecular building blocks of all life discovered as of today. This may seem trite at first, but extrapolating the implications, it enables an even more troubling perspectives on ourselves than simple evolution: (ontological) naturalism. On the lowest level, everything can, in principle, be reduced to elemental particles, their fundamental properties and the reactions among these[2]. One level up, we arrive at complex molecules, crystals and proteins, and the various forms they can take, among which, most importantly to us, is the DNA that defines the genes give rise to life as we know it. Which enables yet another troubling perspective. Seeing (human) life through a gene-centric lens one is forced to not just abandon notions of special creation but also special purpose and meaning. One has to abandon ideas like "People like fruits because they are sweet", and rethink them as "the vehicles (phenotypes) genes built in order to ensure their survival and replication has been provided with a positive reinforcement mechanism for the acquisition of energetically valuable high-fructose nutrition, which it experiences as *sweetness*". Your

[2]One could, of course, go yet another level deeper to quantum mechanics. But as far as we know, nothing that happens there has any tangible relevance for any of arguments presented in this paper. The discussion would just end one level deeper, where everything could be reduced not to particles, but collapsing wave functions.

craving for sweetness exists not because *you like it*, but because, in the evolutionary processes that ended up producing us all, the genes who brought about phenotypes who favour sugary stuff had a competitive edge over those who did not. But because we cannot stop *feeling* special, this is a pretty bitter pill to swallow. To help it go down, maybe the following first instance of robots facilitating philosophical honesty will make the idea more accessible [16, ch.14.4]:

Suppose you wanted to experience life four or five centuries from now. Suppose the only way to accomplish this were to place your body in a cryogenic chamber of sorts and cool down your body (and brain) to almost absolute zero. In this capsule, all your bodily functions are stopped and you will remain suspended in super-coma until re-awoken, as programmed. But the future is unpredictable and dangerous, and as anyone who has seen the popular film *forever young* can easily imagine, someone could accidentally cut off your power supply, a war could be breaking out, or a natural catastrophe could ruin your frozen body's day. One approach to prepare for such unforeseeable circumstances could be to entrust your body to your family and friends. But how sure can you be that over the span of centuries their descendent would continue to maintain such a stewardship? So in order to safeguard of your frozen body, you decide to engineer a super-system around your capsule that will ensure the survival of your body. One option would be a very solid, stable place to settle and build a stronghold there. But even on the toughest rock your capsule might fall victim to all sorts of new problems you simply could not foresee (new anti-gravity motorway maybe?). So the more sophisticated (albeit more expensive) option would be to make your capsule+supporting machinery mobile and thread-responsive, for example in the form of a giant robot equipped with all kinds of sensors and early warning devices, ways to find resources that can be used for energy and self-repair, and an ability to assess and predict the world around it, while you stay safe in your capsule deep inside its guts. You probably see where this is going. Just add some more computing power, maybe the ability to interact and form social groups with other robots and the capsule-in-robot example starts to look like a macroscopic, fully deterministic, fully naturalistic representation of a genes embedded in their biological robot-bodies. If imagined with the current level of robotics, a capsule-robot would probably not exceed the sophistication of a very simple creature. But in principle, there is no reason why one wouldn't be able to imagine a robot refined to such a degree that it could match the complexity and sophistication of a human agent[3]. And if such a level was reached, it would make sense to equip this robot with something akin to consciousness, including preferences that align with your own interests. So, to return to the sweetness of fruit example, clearly there are better and worse ways to acquire energy for your robot and it therefore makes sense for you to program some sort of positive reinforcement mechanism for it to seek out high-yield energy sources. This flow from your intention to keep your cryogenic capsule running, to the interest of the robot to seek out high-yield fuel, is the equivalent of our genes giving us the sensation of pleasurable sweetness when we eat sugary stuff.

Note, however, how this does not eliminate the existence of such phenomena. The point of this elaboration on naturalism, atoms, genes and evolutionary processes is not to deny the existence of our conscious experience of the perception of sweetness. Neither is it a prescription to try reduce everything, even complex human activities and concepts to the level of interacting particles. That would be *greedy reductionism*, the attempt to explain too much with too little [16, p.82]. Clearly, different objects require different levels of abstraction. There would, for example, be little sense in attempting to examine and attempt to explain the cultural phenomenon of sagging pants among teenagers at the level of chemistry or physics. Not because it wouldn't be impossible in principle, but because it would be impossible in practice. So the point is to reach a level of philosophical honesty, where we allow us humans to see ourselves not on a different plane, a different level of existence than our creations (robots), but on exactly the same. The only difference (for now), is that we are mostly made from different materials. There is nothing really *special* about us. We too are walking automatons–just the wet kind, not the dry. So why not try to emphasise this point by changing the terminology that partly defines this differentiation? How about referring to us as bio-bots, not humans and robots as techno-bots? In the following, I will try to do so whenever possible.

3.2 Honest Ethics

Accepting a naturalistic, Darwinian perspective on the world and its agents does not leave ethics untouched. At the very onset, if we take what I outline above seriously and abandon notions of bio-superiority, it calls for yet another round of what has historically been called *the expansion of the circle of ethical consideration* [46]. As far as we can reconstruct from the earliest historic records available, this circle started out with a circumference that only included members of our tribes. In the old testament, for example, it is stated how one may never take tribal members as slaves (even though they may be indebted to you), but may feel free to do so with other people. Similarly, in ancient Greece, only citizens of the respective (city)state possessed rights, while strangers only enjoyed protection from the laws of hospitality (if hosted by a right-bearing citizen). This was still the case when Plato first suggested to expand the circle so that Greeks would no longer fight each other, and only lay waste to foreign nations [39, 469b-71b], [36]. Clearly his suggestion was probably neither the sole nor prime reason, but indeed, in the following centuries the circle of moral concern was expanded to encompass not just tribes and cities but whole

[3]That is, of course, only if you subscribe to a monist perspective on the so called *hard problem of consciousness*. If you believe, as many naturalistically minded neuroscientists and philosophers do [15, 10, 7], that consciousness and the brain are not ontologically distinct entities, that they are ultimately made from atoms, molecules, proteins etc. and their relative properties, and that, eventually, we will be able to explain mental processes in terms of physical terms, then it should not matter whether a brain+mind is made from proteins and neurons or from silicon and transistors. If you do belief in a dualistic interpretation of mind and matter, however, you will probably raise objections at this point. But the burden of proof would clearly be on your side. You would need to explain why, for example, a complete reconstruction of a human brain using different elements would not yield the same results in terms of intelligence, consciousness etc.

nations and states. That is, of course, only the male and racially pure inhabitants of those. Female and differently-skinned individuals were, for the most part, still either not considered morally significant at all or at least significantly inferior as well into the 19th century where it was still considered perfectly normal to, for example, own a black slave or physically abuse a daughter or wife. But starting in the last two centuries, the circle of moral consideration has been expanded drastically. At least on paper, we have seen the circle include bio-bots of different nationalities, race, religion and even gender. Even more recently there have been pushes to also include other bio-bots that come with slightly different genetic make up and phenotype (animals) [45]. The driving force behind these expansions has always been a critical re-evaluation of the attributes that were considered morally significant. First tribal membership was the basis, then nationality, then race, then sex, then species, and, as as of yet: biological origin and consciousness[4].

But if all of these have been succeeded, what then is the final lowest common denominator for moral consideration? Where should the the circle end? With personhood and *the ability to have rights* [40]? The capacity to feel pain or suffer [6]? A particular level of intelligence? The capacity to have and express interests [46]? Each have their proponents and arguments for and against. And while different concepts may be compatible, allow me, in the following, to develop why preference utilitarianism will take the price.

First of all, evolutionary theory allows us to also view morality and ethics from a naturalistic (and less esoteric) perspective. For millennia, ethicists and moral philosophers were convinced that morality must have some sort of independent truth to it, some supernatural origin or essence, something that makes it special and different. As referenced above, still today, 56% of philosophers categorise themselves as moral realists. Which means they too subscribe to a variation of moral exceptionalism, believing in objective moral truths, independent from our perceptions and attitudes towards them, and in a sense akin to the laws of nature[5]. Why is that? Why do they employ this sky hook? Is it not possible to explain morality and ethics with the building bricks supplied by a fully natural, Darwinian world? It appears many believe it is not. They maintain that there must be something else out there, something absolute and different, a special source for good and evil.

Some philosophers and many scientists, however, disagree. One reasons to think otherwise, for example, are the countless observations of moral behaviour that have been made in other animals once we finally started looking. Frans de Waal, for example, has extensively studied the behaviour of non-human bio-bots, and especially that of great apes and monkeys. What one can witness is not just the crude kin altruism one can see in simple bots like bees or ants (who readily sacrifice their lives to save their hives), but also rather refined moral sentiments such as a sense of fairness, justice and so one–patterns of behaviour that had usually been attributed to human bio-bots only. In a highly entertaining study, for example, de Waal offered two capuchin monkey rewards in exchange for a simple task [9]. In the first round, monkey A performs the task and receives a reward, a slice of cucumber, which he gratefully accepts. Monkey B, visible to monkey A performs the same task but receives a grape instead–a snack valued much higher by the capuchin monkeys. Monkey A observes this, repeats the task, and again receives a slice of cucumber. This time, however, he does not gratefully accept the slice and forcefully throws back at the researcher, rattling his cage in visible frustration[6]. This and many similar experiments provide good reason to belief that a sense of what is right and wrong is indeed not unique to us human bio-bots. Evidently, simpler bio-bots are capable of the same, even though they apparently lack the kind of rationality human bio-bots tend to be so proud of. But if not a property of human brains, maybe there is something special about bio-brains in general that allows it to display moral behaviour? Again, from a non-dualist perspective, there is no reason why this would be the case. But for further evidence, consider how, already in 1980, political scientist Robert Axelrod invited game theorists to submit computer programs to a tournament to find out which one would fare best in a multiple-rounds prisoner's dilemma scenario. There were no limitations to the complexity or simplicity of the program (as long as it could still be computed). The strategies ranged from "always defect" to "always cooperate" and "fully random" to elaborate "if this, then that" strategies. To the surprise of the participants, one of the simplest won quite decisively, called TIT FOR TAT and submitted by Anatol Rapoport. Instead of employing complex strategies, it simply always cooperated on the first move and subsequently copied what the opponent's previous move. Since then, many more tournaments of this kind have been hosted and virtually always, the "nice", the programs with generally "kind" initial approaches but retaliatory character won [5, ch.2]. This is significant one the one hand because it provides good reason to believe that what we call morality may have evolved as a successful behavioural strategy for social animals in an evolutionary context, and on the other hand because it provides and examples for even very primitive techno-bot brains being able to be programmed to express moral behaviour.

At this point, however, it is important to clarify that the strategies evolutionary processes likely produced, also do not constitute moral truths or provide ideal moral behaviour. Neither should acknowledging morality's evolutionary roots lead to suggestions of social Darwinism–as many prominent Darwinian scientists and philosophers insist. Darwin himself emphasised to "Never use the words *higher* and *lower*" to guard against possible normative misinterpretations of evolutionary theory [12, p.441]. T.H. Huxley stated that society's ethical progress depends, not on imitation of the cosmic processes, but on combating it [19]. And Richard Dawkins famously wrote: "We are built as gene machines and cultured as meme machines, but we have the power to turn against our creators. We, alone on earth, can rebel

[4]This is where I believe the next expansion will take place when we realise that it does not matter whether you are a bio- or a techno-bot.

[5]There might be a fair amount of semantic bickering about the term "moral realism" that slightly skewed these numbers, but even if corrected, we would probably end up with a pretty high percentage.

[6]If you have not yet seen the video documentation for this experiment, I highly encourage you to look it up right away (https://www.youtube.com/watch?v=meiU6TxysCg).

against the tyranny of the selfish replicators." [13, p.201]. Thus, to believe that "moral" equates to "natural", and to derive a form of social Darwinism would be a case of the naturalistic fallacy [46, p.74], and, without any philosophical justification an attempt to derive an "ought" from an "is", which as David Hume pointed out, simply is not possible within a naturalistic framework [34, p.175].

So, if morality clearly has its origins in evolutionary processes, but at the same time we possess the ability to "rebel against this tyranny", how exactly do human bio-bots conduct their moral business? The answer is: we rely on both; our evolutionary intuitions and our rebelling conscious thoughts about moral situations. As recent works in moral psychology suggest, we tend to conduct our moral reasoning in two connected but distinct modes. Joshua Greene likens these to the different modes of a modern digital camera. On the one hand, one can use the manual mode with lots of options to adjust the lighting, shutter time, focus and so on, representing the conscious, deliberate reasoning about moral problems–or the "rebelling against the tyranny". On the other hand, there is the automatic mode where one has to simply "point and shoot", and the camera takes care of everything else. This represents our tendency to rely on our moral intuitions, much of which find their origins in our evolutionary past for most everyday situations [29, 26]. Jonathan Haidt identifies the same mechanism but prefers a elephant and its rider metaphor to emphasise the imbalance between the two forces [32] (with the elephant representing our moral intuitions and the rider our conscious deliberations). For the most part, the elephant pulls in a particular (moral) direction and we tend to simply follow the pull. And, if asked, we confabulate post-decision rationalisations as we have evolved as elephant-lawyers as Haidt concludes [31].

To illustrate this point, allow me to inject some trolleylology. In the most famous variation of this thought experiment, a bystander faces two moral dilemma scenarios. In situation A, he has to decide whether or not to flip a switch to divert a runaway trolley from track I where a group of five individuals is working to track II with only one individual, and thus whether or not to indirectly kill one to save five. In studies all over the world, the overwhelming majority of participants deems flipping the switch, killing one to save five to be the right decision. This is then followed by a very similar scenario B, but without the switch and only one track. Here, the bystander has to physically push an individual from an overpass to stop the trolley from killing five further down the track. Here, the overwhelming majority of people agree that one should not push the individual–even though the calculation of lives remains exactly the same. Joshua Greene and others take this as prime examples of our two moral modes at work. The first scenario, which requires an abstract action (flipping a switch) engages our manual mode of moral reasoning (or, "the rider" is in charge) and we all agree to the action with the superior consequences. In the second scenario, however, we cannot help but engage in automatic mode (or, "the elephant" is in charge). And even though, when asked, we confabulate reasons why pushing the man is wrong, when pressed for answers we admit that it simply *feels* wrong. The most probable explanation for this an innate moral intuition and bias against causing direct physical harm to undeserving individuals–a valuable innate behavioural pattern to have for social bio-bots[7].

In a sense, all this can be seen as an extension of the Freudian revolution as despite us *feeling* like we are in control and in manual mode, we really are not. The troubling thing is that it appears that very little of contemporary moral philosophy recognise these facts. And many still treat our moral intuitions as self-evident moral axioms[8] and/or our special way to access some sort of independent moral truths. So what are we to do with situation? When much of moral theory depends on what "may be no more than a relic of our evolutionary history" [46, p.70]? Does it make sense to continue the use of such foundations, despite our knowledge about their messiness? And should we attempt to infuse techno-bots with a simulation of this? Should a techno-bot be programmed to flip the switch but not push the person, mirroring the results from the trolley-studies outlined above? I believe not. Instead, what is needed for both bio- and techno-bots is a new understanding of ethics [46, ch.6], which Greene calls morality^2, an objective framework derived from reason and deliberation alone, and workable irrespective of what moral tribe may have been brought up in [28, ch.8]; a rigorous, deliberate manual mode that is well aware of the elephant but refuses to let go of the reigns[32]; An ethics informed by science, aware of our cognitive biases and moral intuitions but judging each situation as objectively as possible [46, ch.6]. Greene and Singer both conclude that a form of consequentialism–in Singer's case preference-utilitarianism, and in Greene's case what he calls "deep pragmatism", a form of classic well-being based utilitarianism–is best fit for this task, as, by their very definitions, they leaves aside everything but the rational calculation of interests/well-being. But regardless of which framework ultimately takes the prize for best fit for morality^2, it is clear that, at least for now, human bio-bots will not be able to fully implement any such morality^2 framework–especially not in their daily lives. Nevertheless, it should be the standard we should strive to approximate.

3.3 Honest Agency

Reflecting on the arguments above, we now have arrived at a level of philosophical honesty where we can realise ourselves as mere bio-bots, a lot more developed, but in no way better or more special than techno-bots. We have seen how our own morality finds its roots in evolutionarily advantageous patterns of behaviour and certain cultural priming and how we tend to use two different modes of moral reasoning whenever we encounter moral situations. One could say this is already plenty of useful material for the task of making robots ethical. But there is one more extension of the Freudian revolution that needs to be discussed before we can conclude with particular prescriptions. It concerns one of our most important intuitions: our deep conviction that

[7]While the results of this particular experiment appear to be culture-independent, there are other examples where not innate (biological) but cultural (or "memetic"to use Dawkin's or Dennett's terms) seems to be the driving force behind the automatic mode of moral judgement (or, one could say a cultural elephant was pulling, not a genetic one). This does not refute the dual-process theory, but simply adds another level [29, ch.4, 10], [25].

[8]Much of deontological thought, for example, is based on moral intuitions (which may not be shared by other moral tribes) [29, ch.7].

we possess free will, that we are the sole causers and fully in charge of our decisions.

As Darwinian revolution taught us, us bio-bots too are constructed from atoms, molecules, proteins and so on. And if, as naturalism claims, there is nothing outside these fundamental particles and their respective properties, and these are the building blocks for the brains that produce our minds, then how could any decision made by such a mind still be "free" and independent? As Daniel Wegner argues in *The illusion of conscious will* [49], the reason why we believe to be exempt from what we otherwise readily describe as a deterministic universe, is unawareness of the processes that are going on in our brains. To us, it appears that our actions are indeed caused by the mental states we observe and not by the physical states in our brains. We strongly *feel* like we are metaphysically special, and that we truly are in charge of what we do. This intuition and final extension of the Freudian revolution is so hard to concede that philosophers engage in all kinds of complex conceptual contortions and employ all sorts of sky hooks just to save at least some notion of free will (which, to a certain degree includes Dan Dennett himself[17]).

To illustrate the issue, let us return to the capsule-in-robot example introduced above. Suppose the bot we constructed to protect our bio-bot body in cryogenic sleep encounters a situation that resembles the trolley dilemma outlined above. A driverless trolley is hurdling down the tracks, unstoppable and on its way to kill five innocent human bio-bots. In this situation, the robot could step in the way of the trolley and bring it to a halt before it reaches the five human bio-bots. The problem is, by doing so, it calculates a significant risk to jeopardise the survival chances of the capsule he carries. And as it has been programmed to place the safety of the capsule above everything else, it turns away and leaves the five human bio-bots to die. Would you consider the bot morally responsible? I expect you probably don't. But what if the bot's decision was not made by a programmed computer, but by a human bio-bot pilot, operating the bot by remote? Imagine the same situation and decision to let the five bio-bots die in order to save the one inside the bot. Would you consider this pilot morally responsible? Can he/she be made morally accountable for his actions and decisions? If you believe he/she can, you too have fallen victim of this potent intuition, this most notorious notion of human specialness.

In their widely cited paper *For the law, neuroscience changes nothing and everything* Joshua Greene and Jonathan Cohen provide a powerful though experiment, aiming to aid you to overcome (at least cognitively) the free will intuition. Based on some recent philosophical discussions on the matter [41], they introduce a thought experiment where a group of mad scientists try to design a very bad, violent person by maintaining tight controls over his genetic make up (possibly taken from known criminals) and experiences during his upbringing (simulating a very troubled childhood) and succeeded in their efforts. As designed, he engages in all sorts of crimes and violence. Eventually the individual is caught and brought before court, where the defence calls to the stand one of the scientists who helped to create him and asked about his role in the matter. He states:

"It is very simple, really. I designed him. I carefully selected every gene in his body and carefully scripted every significant event in his life so that he would become precisely what he is today. I selected his mother knowing that she would let him cry for hours and hours before picking him up. I carefully selected each of his relatives, teachers, friends, enemies, etc. and told them exactly what to say to him and how to treat him. Things generally went as planned, but not always. For example, the angry letters written to his dead father were not supposed to appear until he was fourteen, but by the end of his thirteenth year he had already written four of them. In retrospect I think this was because of a handful of substitutions I made to his eighth chromosome. At any rate, my plans for him succeeded, as they have for 95% of the people I've designed. I assure you that the accused deserves none of the credit." [27, p.1780].

Insofar as one believes the testimony, Greene and Cohen argue, it would be hard to find good arguments in favour of holding this bio-bot responsible for his actions. Yes, it may make sense to describe him as a bad person and one to stay away from, but given the degree to which the scientists maintained control, it is hard to see him as anything but a pawn. Which, of course, raises the question how the rest of us bio-bots are any different from this Mr. Puppet. One may respond that he was obviously part of a diabolical plot, while the rest of us generally lacks such evil puppeteers. But does the evil plot and the scientists' intentions really matter? Could we consider Mr. Puppet to have act out of free will and able to be held morally responsible had he been "designed" to be a good-doing hero by some no less mad, but slightly nicer scientists? I think not. The only thing that matters is that the forces (his genetic make up, his environment and experiences) that made and make Mr. Puppet who he was and is, were, and continue to be beyond his control. Were exactly the same forces present by mere chance, not careful design, the Mr. Puppet they would produce would be no less determined. And, of course, if this is true for Mr. Puppet it is also true for any other bio-bot, including you and me.

Seen through our humans are bio-bots lens, what the mad scientists did was the active programming of a particular bio-bot. And in that sense, there is no difference between the bot and the pilot deciding not to step on the tracks. The only difference is *how* these two bots were programmed. One had its behaviour pattern conveniently uploaded directly to its techno-bot "brain", the other received it through a mix of genetic predisposition and a lifetime of experiences and influences on its bio-bot brain. Thinking this way necessitates to gets rid of moral agency as a concept for an entity could be held morally responsible for the actions that flow from his/her presumed free will.

This, of course, goes very much against how we feel the world works. We may have accepted that the world isn't flat, that it is indeed orbiting the sun and that we may share common ancestors with the animals around us, but it seems almost impossible to accept the thought that there may not be freedom of will. It seems to be so clearly visible in our conscious experience that it just *feels* wrong to suggest it wouldn't exist. This innateness of the feeling suggests that we might be dealing with something similar to our moral intuition not

to push the individual off the foot bridge, that the notion of free will may have similar evolutionary origins. And sure enough, there are some who argue that seen as an evolutionary biological feature, it may have its roots in functioning as a generator of pseudo-randomness, as bio-bots equipped with this feature would be able to decrease the predictability of their actions and thus increase their chances of survival [11]. On a higher level, seen as a part of our innate folk psychology, it is serves as a way to deal with other assumed agents. In general, the argument goes, us bio-bots classify things into two major entities: inanimate stuff, things that just sit there and generally do nothing–rocks, branches of a tree, etc.– and things that move about and do things. The former can easily be predicted and dealt with. But the latter is much more complex and often comes with a much higher significance for our lives. Because these two categories of things are so fundamentally different in their importance to us to process them with two different cognitive modes (Dan Dennett calls them the natural and intentional stance respectively [14]) makes sense from an evolutionary perspective [38]. While we experience and understand things like stones and branches etc. from a very physical point of view (If I drop a stone, it falls to the ground), we use the concepts of agency, intentions, feelings and so on to describe and predict the behaviour of things that move about and do things (If I drop a cat, it will turn around because it *wants to* land on its feet). Many experiments have provided evidence for our innate urges to ascribe these concepts even to the most primitive things. In the famous experiments run by Heider and Simmel in 1944, for example, the movement of a simple triangle and circle was repeatedly interpreted in social terms, with one described as *threatening*, *bullying* or *trying to protect* the other respectively [33]. These responses appear to be completely automatic and cannot be turned off [43, 2]–that is unless you critically damage a particular part of your amygdala, in which case you may be able to describe the actions and movements of the circles and triangles in completely abstract and asocial terms [1]. In other words, we cannot help but to ascribe other animate things the same properties we experience for ourselves: the conscious presence of a mind capable of intentions, desires and acting upon free will–which appears to be at the very heart of what it takes for us to consider something a (moral) agent [43]. And because we perceive ourselves as autonomous, un-caused agents in possession of free will, we attribute the same to others and feel justified in assigning blameworthiness and praiseworthiness for their and our own actions, intensifying these judgements with the level of intentionality we detect [47]. Like our moral intuition not to push the individual onto the tracks, our sense of possessing free will appears to be a remnant of our evolutionary past, useful for our more primitive ancestors, but today obsolete as we learn about its origins and are capable to "rebel against its tyranny".

So, where to go from here? If there is no free will, doesn't this also dismiss ethics all together? Well, yes and no. Yes, because it does indeed eliminate all retributive notions–the kind of thinking where we believe that one should "get what one deserves". No, because, for the consequentialist approach (and especially the preference-utilitarian one), for example, nothing of significance changes. Objectively, any action can still cause harm or do good, regardless of whether the robot (bio- or techno-) has performed it out of "free will" or due to deterministic forces. Which is why, from a legal point of view grounded in utilitarian principles, everything and nothing changes at the same time as Greene and Cohen conclude in their paper [27]. We cannot blame him for it, but Mr. Puppet really does commit crimes and does real harm to other people. Hence it still may be necessary to lock him up in an attempt to "reprogram" him to be a better citizen, and if this fails, we may even have to lock him up for life to protect everyone else. In a redefinition of what it means to be a moral agent, we can hold him accountable, but not morally responsible [21].

4. ROBOT ETHICS

Having followed the Darwinian and Freudian revolution to their advanced conclusions, we then have arrived at a level of philosophical honesty I believe could be quite helpful for the task of making techno-bots ethical. As a by-product, we have also arrived at consequentialism as an ethical framework compatible with this honesty. This is not to say it is the only one to do so and not a definitive prescription. It simply serves as an example how one need not abandon ethics all together when embracing a fully naturalistic, deterministic, "honest" world-view–in fact, quite the opposite, it makes ethics clearer and less messy, and quite possibly much more applicable for robots.

4.1 Ontological Status

As an extension of the Darwinian evolution allows us to conclude, bio-bots and techno-bots are not members of different ontological categories. From an objective (naturalistic) perspective, both are potentially intelligent (intentional) entities in a fully deterministic system. They do differ in the material they are made of, but this is of no significance for their ability to possess "brains" capable of thinking, having preferences or acting as moral agents and patients within a morality[2] framework.

4.2 Moral Patiency

At least within a utilitarian framework, it also does not matter how a bot of any kind acquired its preferences/sense of well-being. Whether these are the result of genes and experiences, or of a programme uploaded to a brain-like functioning computer, we ought to recognise them the same (this is the next expansion of the circle of moral consideration as outlined above). That said, especially with regards to the simple bots we already see today, it does matter whether they possess the preference to have their preferences/well-being recognised. This may seem like a cheap excuse to treat less complex bots poorly, but I believe it to be of relevance. After all, it appears to be the underlying premise of why we care about our own preferences/well-being and that of others in the first place: because we care about our own and extrapolate similar meta-preferences for others. If you have a simple techno-bot programmed with only one "interest"–to seek out the dark spot of a room–you ought to take this interest into consideration. At the same time, however, there is nothing objectionable about denying it its interest as it does not possess the meta-preference of not having its preferences denied. Equally, if you do not program a bot with a sense of well-being, there is no reason for moral consideration of what isn't present. At the current level of AI

research, I doubt any robot would qualify to have such preferences or sense of well-being. In the future, however, it is very much thinkable that AI may develop such properties by itself or have it bestowed by a bio-bot creator. As soon as this may happen, we may no longer easily disregard the preferences/well-being of such techno-bots and place equal weight to the consideration of them as we do for ours or that of other bio-bots [42].

4.3 Moral Agency

As moving objects with the potential for complex behaviour that might interfere with the preferences and well-being of other such objects (bio-bots and techno-bots), techno-bots are as much capable of ethical or unethical behaviour as we are, and thus qualify as moral agents. Not in the retributive sense as discussed above, but in the objective, independent view of morality^2/consequentialism. It allows us to hold fully deterministic entities not morally responsible, but ethically accountable and responds with careful reprogramming to protect and increase net preference/well-being-maximisation. And as techno-bots are (for now), considerably easier to reprogram than bio-bots, and because we have the opportunity to provide them with morality^2 right from the start, they may soon be better moral agents than us bio-bots are as of yet (with our messy elephant+rider morality).

5. OUTLOOK

For a naturalistic philosopher, none of the above should be particularly surprising or new. Nevertheless, I believe it is worth to connect these perspectives and ways of thinking to the case of robot ethics. Techno-bots will undoubtedly soon be as much part of our lives as companion bio-bots (pets) and computers already are [24]. For the foreseeable future, they will most likely remain quite simple mechanical slaves. But there is good reason to believe there will be a point in the future where strong artificial intelligence will be possible, and our robot slaves will gain "consciousness". When that happens, we can only hope to have settled for an ethical framework that is fit to deal with techno- and bio-bots alike, for our sake and theirs. This paper represents the attempt to contribute to this task.

6. REFERENCES

[1] R. Adolphs. Social cognition and the human brain. *Trends in Cognitive Sciences*, 3(12):469–479, Dec. 1999.

[2] R. Adolphs. How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, 1079(1):25–35, Mar. 2006.

[3] C. Allen, W. Wallach, and I. Smit. Why Machine Ethics? In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 51–60. Cambridge University Press, New York, May 2011.

[4] S. L. Anderson. How Machines Might Help Us Achieve Breakthrougs in Ethical Theory and Inspire Us to Behave Better. In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 151–160. Cambridge University Press, New York, May 2011.

[5] R. Axelrod. *The Evolution of Cooperation: Revised Edition*. Basic Books, Dec. 2006.

[6] J. Bentham. *An Introduction to the Principles of Morals and Legislation*. Clarendon Press, Oxford, 1823. bibtex: bentham_introduction_1823.

[7] S. Blackmore. *Conversations on Consciousness: What the Best Minds Think about the Brain, Free Will, and What It Means to Be Human*. Oxford University Press, New York, 1 edition edition, Jan. 2007.

[8] D. Bourget and D. J. Chalmers. What Do Philosophers Believe? *Philosophical Studies*, 170:465–500, Nov. 2013.

[9] S. F. Brosnan and F. B. M. de Waal. Monkeys reject unequal pay. *Nature*, 425(6955):297–299, Sept. 2003.

[10] P. M. Churchland. *Matter and Consciousness*. The MIT Press, Cambridge, Massachusetts, third edition edition edition, Aug. 2013.

[11] L. F. Clegg. Protean Free Will. Tufts University, 2012.

[12] C. Darwin and J. T. Costa. *The Annotated Origin: A Facsimile of the First Edition of On the Origin of Species*. Harvard University Press, 2009.

[13] R. Dawkins. *The Selfish Gene: 30th Anniversary Edition–with a new Introduction by the Author*. Oxford University Press, Oxford ; New York, 30th anniversary edition edition, May 2006.

[14] D. C. Dennett. *The Intentional Stance*. A Bradford Book, Cambridge, Mass., reprint edition edition, Mar. 1989.

[15] D. C. Dennett. *Consciousness Explained*. Back Bay Books, Boston, 1 edition edition, Oct. 1992.

[16] D. C. Dennett. *Darwin's dangeorus idea: evolution and the meanings of life*. Simon & Schuster Pperbacks, New York, 1995.

[17] D. C. Dennett. *Freedom evolves*. Penguin UK, 2004.

[18] D. C. Dennett. Computers as prostheses for the imagination. Laval, France, 2006.

[19] J. Dewey. Evolution and Ethics. *Monist*, VIII:321–341, 1898.

[20] A. Ellegard. *Darwin and the General Reader: The Reception of Darwin's Theory of Evolution in the British Periodical Press, 1859-1872*. University of Chicago Press, 1958.

[21] L. Floridi. Artificial Intelligence's New Frontier: Artificial Companions and the Fourth Revolution. *Metaphilosophy*, 39(4-5):651–655, Oct. 2008.

[22] L. Floridi. On the Morality of Artificial Agents. In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 151–160. Cambridge University Press, New York, May 2011.

[23] L. Floridi. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Also available as: eBook, June 2014.

[24] B. Gates. A Robot in Every Home. *Scientific American*, 296(1):58–65, 2007.

[25] J. Graham, J. Haidt, and B. A. Nosek. Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046, 2009.

[26] J. Greene. Beyond point-and-shoot morality: why cognitive (neuro) science matters for ethics. *Ethics*, 124(4):695–726, 2014.

[27] J. Greene and J. Cohen. For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 359(1451):1775–1785, Nov. 2004.

[28] J. D. Greene. *The Terrible, Horrible, No Good, Very Bad Truth About Morality and What To Do About It*. Dissertation, Department of Philosophy, Princeton University, Princeton, N.J., 2002.

[29] J. D. Greene. *Moral tribes: emotion, reason, and the gap between us and them*. The Penguin Press, New York, 2013.

[30] D. J. Gunkel. *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*. The MIT Press, Cambridge, Mass, July 2012.

[31] J. Haidt. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4):814–834, 2001.

[32] J. Haidt. *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Vintage, New York, reprint edition edition, Feb. 2013.

[33] F. Heider and M. Simmel. An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, 57(2):243–259, Apr. 1944.

[34] D. Hume. *A Treatise of Human Nature*. Clarendon Press, Oxford, 1739. bibtex: hume_treatise_1739.

[35] D. G. Johnson. Computer Systems – Moral Entities but Not Moral Agents. In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 151–160. Cambridge University Press, New York, May 2011.

[36] R. Kamtekar. Distinction Without a Difference? Race and GEnos in Plato. In J. K. Ward and T. L. Lott, editors, *Philosophers on Race: Critical Essays*. John Wiley & Sons, June 2008.

[37] F. Newport. In U.S., 42% Believe Creationist View of Human Origins. Technical report, Gallup, June 2014.

[38] S. Pinker and M. Foster. *How the Mind Works*. Brilliance Audio, mp3 una edition edition, Apr. 2014.

[39] Plato and J. M. Cooper. *Plato: Complete Works*. Hackett Publishing Co., Indianapolis, Ind, May 1997.

[40] T. Regan. *The Case for Animal Rights*. University of California Press, 1985. bibtex: regan_case_1985.

[41] G. Rosen. The Case for Incompatibilism. *Philosophy and Phenomenological Research*, 64(3):699–706, May 2002.

[42] A. Sagan and P. Singer. Rights for Robots? *Project Syndicate*, Dec. 2009.

[43] B. J. Scholl and P. D. Tremoulet. Perceptual causality and animacy. *Trends in Cognitive Sciences*, 4(8):299–309, Aug. 2000.

[44] H. Seville and D. G. Field. What Can AI Do for Ethics? In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 499–511. Cambridge University Press, New York, May 2011.

[45] P. Singer. *Animal liberation: A new ethics for our treatment of animals*. New York Review, New York, 1975. bibtex: singer_animal_1975.

[46] P. Singer. *The expanding circle: Ethics, evolution, and moral progress*. Princeton University Press, 2011. bibtex: singer2011expanding.

[47] G. Strawson. On "Freedom and Resentment". In J. M. Fischer, editor, *Free Will: Concepts and challenges*. Taylor & Francis, 2005.

[48] J. P. Sullins. When Is a Robot a Moral Agent? In M. Anderson and S. L. Anderson, editors, *Machine Ethics*, pages 151–160. Cambridge University Press, New York, May 2011.

[49] D. M. Wegner. *The Illusion of Conscious Will*. A Bradford Book, Cambridge, Mass., 1 edition edition, Aug. 2003.

[50] R. S. Westman. *The Copernican Question: Prognostication, Skepticism, and Celestial Order*. University of California Press, Berkeley, July 2011.

Robots, ethics and language

Ingrid Björk
Uppsala University
P.O.Box 337
SE-751 05 Uppsala, Sweden
+46736165961
Ingrid.Bjork@lingfil.uu.se

Iordanis Kavathatzopoulos
Uppsala University
P.O. Box 337
SE-751 05 Uppsala, Sweden
+46704250383
iordanis@it.uu.se

ABSTRACT
Following the classical philosophical definition of ethics and the psychological research on problem solving and decision making, the issue of ethics becomes concrete and opens up the way for the creation of IT systems that can support handling of moral problems. Also in a sense that is similar to the way humans handle their moral problems. The processes of communicating information and receiving instructions are linguistic by nature. Moreover, autonomous and heteronomous ethical thinking is expressed by way of language use. Indeed, the way we think ethically is not only linguistically mediated but linguistically construed – whether we think for example in terms of conviction and certainty (meaning heteronomy) or in terms of questioning and inquiry (meaning autonomy). A thorough analysis of the language that is used in these processes is therefore of vital importance for the development of the above mentioned tools and methods. Given that we have a clear definition based on philosophical theories and on research on human decision-making and linguistics, we can create and apply systems that can handle ethical issues. Such systems will help us to design robots and to prescribe their actions, to communicate and cooperate with them, to control the moral aspects of robots' actions in real life applications, and to create embedded systems that allow continuous learning and adaptation.

Categories and Subject Descriptors
K.4.1 [**Public Policy Issues**]: Ethics.

General Terms
Design, Security, Human Factors, Languages, Theory.

Keywords
Autonomy, communication, decision-making, design, ethics, independent agents, language, moral, philosophy, robots.

1. INTRODUCTION
Automated IT systems can be of great help to achieve goals and obtain optimal solutions to problems in situations where humans have difficulties to perceive and process information, or make decisions and implement actions because of the quantity, variation and complexity of information. One example, relating to human social and emotional needs, is elderly care where robots may come to play an important role, in providing necessary care as well as in supplying continuous stimulation to lonely elderly people.

It is clear that automated IT systems have to make decisions and act to achieve the goals for which they had been built in the first place but there are many questions to address around the issue. First, will they make the right decisions and act in a proper way? Second, can we guarantee that they do by designing them in a suitable way? Third, even if we can control their actions, do we really want such constrained machines, given the fact that the main reason we want them in the first place is their increasing independence and autonomy?

Most of the questions converge on the issue of moral or ethical decision making. The definition of what we mean by ethical or moral decision making or ethical/moral agency is a very significant precondition for the design of proper automated decision systems. Given that we have a clear definition based on philosophical theories and on research on human decision-making, we want to create and apply systems that can handle ethical issues of independent agents. Such systems will help us to design agents and to prescribe their actions, to help us control the moral aspects of agents' actions in real life applications, and to create embedded systems that allow continuous learning and adaptation.

2. ETHICS
The distinction between content and process is important in the effort to define ethical or moral decision-making**.** In common sense, ethics and morals are dependent on the concrete decision or the action itself. Understanding a decision or an action being ethical/moral or unethical/immoral is based mainly on a judgment of its normative qualities. The focus on values and their normative aspects is the basis of the common sense definition of ethics. For example, it is supposed that independent military robots have to follow the laws of war to be called ethical [2].

Despite its dominance, this way of thinking causes some difficulties. We may note that bad or good things follow not only from the decisions of people but also from natural phenomena. Usually sunny weather is considered a good thing, while rainy weather is not. Of course this is not perceived as something related to morality. But why not? What is the difference between humans and nature acting in certain ways? The answer is obvious: Option, choice.

Although common sense does realize that, our attachment to the normative aspects is so strong that it is almost impossible to accept that ethics is an issue of choice and option. If there is no choice, or ability of making a choice, then there is no issue of ethics. However this does not solve our problem of the definition

of what an independent ethical agent would be, since many IT systems are actually making choices.

If ethics is connected to choice then the interesting aspect is how the choice is made, or not made; whether it is made in a bad or in a good way. The focus here is on *how*, not on *what*; on the process not on the content. Indeed, regarding the effort to make the right decision, philosophy and psychology point to the significance of focusing on the process of ethical decision making rather on the normative content of the decision.

Starting from one of the most important contributions, the Socratic dialog, we see that *aporia* is the goal rather than the achievement of a solution to the problem investigated. Reaching a state of no knowledge, that is, throwing aside false ideas, opens up for the right solution. The issue here for the philosopher is not to provide a ready answer but to help the other person in the dialog to think in the right way [17], [19]. Ability to think in the right way is not easy and apparently has been supposed to be the privilege of the few able ones [18]. For that, certain skills are necessary, such as Aristoteles's *phronesis* [1]. When humans are free from false illusions and have the necessary skills they can use the right method to find the right solution to their moral problems [5].

This classical philosophical position, supported also by modern philosophers like Popper, Habermas, Foucault, Sartre and others, has been applied in psychological research on ethical decision-making. Focusing on the process of ethical decision-making psychological research has shown that people use different ways to handle moral problems. According to Piaget [16] and Kohlberg [14], when people are confronted with moral problems they think in a way which can be described as a position on the heteronomy-autonomy dimension. *Heteronomous* thinking is automatic, emotional and uncontrolled thinking or simple reflexes that are fixed dogmatically on general moral principles. Thoughts and beliefs coming to mind are never doubted. There is no effort to create a holistic picture of all relevant and conflicting values in the moral problem one is confronted with. Awareness of own personal responsibility for the way one is thinking or for the consequences of the decision is missing. *Autonomous* thinking, on the other hand, focuses on the actual moral problem situation, and its main effort is to search for all relevant aspects of the problem. When one is thinking autonomously the focus is on the consideration and investigation of all stakeholders' moral feelings, duties and interests, as well as all possible alternative ways of action. In that sense autonomy is a systematic, holistic and self-critical way of handling a moral problem.

Handling moral problems autonomously means that a decision maker is unconstrained by fixations, authorities, uncontrolled or automatic thoughts and reactions. It is the ability to start the thought process of considering and analyzing critically and systematically all relevant values in a moral problem situation. This may sound trivial, since everybody would agree that it is exactly what one is expected to do in confronting a moral problem. But it is not so easy to use the autonomous skill in real situations. Psychological research has shown that plenty of time and certain conditions are demanded before people can acquire and use the ethical ability of autonomy [20].

Nevertheless, there are people who have learnt to use autonomy more often, usually people at higher organizational levels or people with higher responsibility [12], [13]. Training and special tools do also support the acquisition of the skill of autonomy. Research has shown that it is possible to promote autonomy. It is possible through training to acquire and use the skill of ethical autonomy, longitudinally and in real life [6], [7], [8].

However, ethical competence is not the use of autonomy every time a moral problem has to be solved. Rather, it is the ability to use it if and when the problem at hand demands it; not to use it always and for any kind of moral problem. On the other hand, heteronomy is actually working, despite the fact being an automatic, mostly unconscious and a constrained way to handle moral problems. People use it most of the time and they repeatedly manage to produce satisfactory solutions to their problems. Furthermore, people facing a moral problem do not adopt purely autonomous or heteronomous ways of handling it in their efforts to solve it and to make a decision. They use a mix of these two ways. And most often they adopt ways that are dominated by heteronomy [12], [13].

Why is that then? Well, the obvious answer is that ways dominated by heteronomy in fact lead to the achievement of decision makers' mail goals. People's goal is not to use perfect ways to solve their moral problems. They just want to get satisfactory solutions to what they feel is important for them. In that sense it is highly significant for them to avoid uncertainty and anxiety as well as big investments in resources, effort and time that follow with autonomy.

3. COMMUNICATING

All this means that we can create working tools to support ethical problem solving and decision-making. This is important since ethics is generally perceived to be too theoretical to be applicable in practice, especially regarding the area of IT. However, following the classical definition of ethics and the psychological research on problem solving and decision-making, the issue of ethics becomes concrete and opens up the way for the creation of IT systems that can support handling of moral problems. Also in a sense that is similar to the way humans handle their moral problems.

The processes of communicating information and receiving instructions are linguistic by nature. Moreover, autonomous and heteronomous thinking is expressed by way of language use. Indeed, the way we think ethically is not only linguistically mediated but linguistically construed – whether we think for example in terms of conviction and certainty (meaning heteronomy) or in terms of questioning and inquiry (meaning autonomy). A thorough analysis of the language that is used in these processes is therefore of vital importance for the development of the above mentioned tools and methods. For this purpose an analysis based on the theoretical framework of Critical Discourse Analysis (CDA) would be suitable.

Critical Discourse analysis has been used in sociology by e.g. Norman Fairclough. Fairclough's model includes three inter-related dimensions of discourse, *text, socio-cultural practices* and *discursive practices*, each of which requires a different kind of analysis [3]. Fairclough regards text, (linguistic utterance), discursive practice (processes by which text is created and consumed) and socio-cultural practice as both constituting and constituted by each other and examines in his research how language figures as an element in social change, and the relation between language use and societal, socio-cultural patterns.

Research in CDA often draws on the theory of Systemic Functional Linguistics (SFL) [4] which focuses on the function of linguistic categories and analyses language as shaped by the social functions it serves. SFL is used for linguistic analysis on lexico-grammatical level as well as on higher text levels, and sets up a relationship all the way from the very concrete words and grammatical structures to the more abstract levels of context such as ideology. In this way, particular linguistic features of a text such as words and grammar may be related to ideology, attitude and intent.

4. ROBOTS AND ETHICS

There are three major avenues for incorporating ethical decision making in automated systems. We will outline them here, ordered by increasing technical challenge. The first and most straightforward is to perform an analysis to determine the impact of the system on society and people's interests.

We have constructed and worked with different versions of *EthXpert/ColLab/Democrate*, http://www.it.uu.se/research/project/ethcomp/ethxpert/, which are tools intended to support the process of structuring and assembling information about situations with possible moral implications. Analogous with the deliberation of philosophers throughout history as well as with the findings of psychological research on ethical decision-making, *EthXpert/ColLab/Democrate* has been built on the hypothesis that moral problems are best understood through the identification of authentic interests, needs and values of the stakeholders in the situation at hand. Since the definition of what constitutes an ethical decision cannot be assumed to be at a fix point, we further conclude that this kind of system must be designed so that it does not make any assertions of the normative correctness in any decisions or statements. Consequently, the system does not make decisions and its sole purpose is to support the decision maker (a person, a group or an organization) when analyzing, structuring and reviewing choice situations [10], [11], [15].

Tools like *EthXpert/ColLab/Democrate* can be used during the development of agents or decision-making systems to identify the criteria for making decisions and for choosing a certain direction of action to be programmed into the agent prescribing how it will act. This means that the support tool is used by the developers; the ones who make the real decisions and thus should make them according to the previously mentioned philosophical and psychological position. In this case designers get help by *EthXpert/ColLab/Democrate* to use autonomy whereas the agents follow their instructions in a heteronomous way.

The second approach is slightly more challenging, yet still perfectly manageable. Based on the same theoretical framework as the first approach, the idea is to implement an ethical decision preparation system in the automated system. Following an initial over-enthusiasm about automation, the recent decades of experience have led monitoring systems to evolve, from delivering emergency data to a passive recipient, to involve the operator more in the decision making. The reason for this is obvious: active operators become better at handling also infrequent problems, simply because they have more experience and continuously exercise their expertise. Current systems are often good at relieving cognitive stress but we should not be satisfied with this. A future research direction should be aimed at creating systems that help operators to take in account systemic features, like environmental and human values. This suggests reconsiderations about the way that an operator is being presented with data from the system and the linguistic character of the information. One example would be about robot assistants in health care, where the feedback to operators needs to comprise not only technical details but also operationalizations of human values.

Communication and language use is clearly an important issue here. Language is not only a vehicle for transferring information but has potential to hinder or stimulate the autonomous way of thinking. The linguistic dimension is significant in the efforts to support the right way of ethical thinking.

This means that we integrate a support tool, like *EthXpert/ColLab/Democrate*, into the agent or the decision system. Of course, designers can give to the system criteria and directions, but they can also add the support tool itself, to be used in the case of unanticipated future situations. The tool can then gather information, treat it, structure it and present it to the operators of the decision system in a way which follows the requirements of the above mentioned theories of autonomy. If it works like that, the operators of agent systems will still be in charge of making the real decisions as they are the users of the ethical support tool. A system like that can make decisions and act in accordance to the hypothesis of ethical autonomy by having the criteria already programmed in it identified through an autonomous way in an earlier phase by the designers. Later on, when in action, the agent by the help of a tool like *EthXpert* can gather and prepare the information of a problem situation, present it and stimulate the operator to make the decision in an autonomous way, compatible with the above mentioned philosophical and psychological theories.

All this can work and it is possible technically. But how could we design and run a really independent ethical decision making system? The third and final step means to implement automatic judgment in trained autonomous systems. A possible way is to design self-learning agents by the use of a tool like *EthXpert/ColLab/Democrate* balanced by a flexible system of blocking. But this is an avenue with several both theoretical and practical obstacles [9], [11], [21]. However, as many systems require complex and faster decision making than is possible for humans it may be an unavoidable development. In these kinds of systems it is crucial to create predictability and traceability.

5. REFERENCES

[1] ’Αριστοτέλης [Aristoteles]. 1975. *’Ηθικά Νικομάχεια [Nicomachean Ethics]*. (Πάπυρος [Papyros], Αθήνα [Athens]).

[2] Arkin, R.C. 2009. *Governing lethal behavior in autonomous robots*. Taylor & Francis Group, Boca Raton FL.

[3] Fairclough, N. 1995. *Media discourse*. London. Hodder Arnold.

[4] Halliday, M.A.K. 1994. *An introduction to functional grammar*, 2nd edn. London: Edward Arnold.

[5] Kant, I. 2006. *Grundläggning av sedernas metafysik* [Groundwork of the metaphysic of morals]. Daidalos, Stockholm.

[6] Kavathatzopoulos, I. 1993. Development of a cognitive skill in solving business ethics problems: The effect of instruction. *Journal of Business Ethics*, 12, 379-386.

[7] Kavathatzopoulos, I. 1994. Training professional managers in decision-making about real life business ethics problems: The acquisition of the autonomous problem-solving skill. *Journal of Business Ethics*, 13, 379-386.

[8] Kavathatzopoulos, I. 2004. Making ethical decisions in professional life. In H. Montgomery, R. Lipshitz and B. Brehmer (Eds.), *How professionals make decisions* (pp. 277-288). Lawrence Erlbaum Associates Inc., Mahwah, NJ.

[9] Kavathatzopoulos, I. 2014. Independent agents and ethics. In K. Kimppa (Ed.), *ICT and society* (pp. 39-46). Springer, Heidelberg.

[10] Kavathatzopoulos, I. and Laaksoharju, M. 2010. Computer aided ethical systems design. In M. Arias-Oliva et al. (Eds.), *The "backwards, forwards, and sideways" changes of ICT* (pp. 332-340). Universitat Rovira i Virgili, Tarragona, Spain.

[11] Kavathatzopoulos, I. and Laaksoharju, M. 2011. What are ethical agents and how can we make them work properly? In C. Ess and R. Hagengruber (Eds.), *The computational turn: Past, presetns, futures?* (pp.151-153). Münster: MV-Wissenschaft.

[12] Kavathatzopoulos, I. and Rigas, G. 1998. A Piagetian scale for the measurement of ethical competence in politics. *Educational and Psychological Measurement*, 58, 791-803.

[13] Kavathatzopoulos, I. and Rigas, G. 2006. A measurement model for ethical competence in business. *Journal of Business Ethics Education*, 3, 55-74.

[14] Kohlberg. L. 1985. The Just Community: Approach to moral education in theory and practice. In M. Berkowitz and F. Oser (Eds.), *Moral education: Theory and application* (pp. 27-87). Lawrence Erlbaum Associates, Hillsdale NJ.

[15] Laaksoharju, M. and Kavathatzopoulos, I. 2009. Computerized support for ethical analysis. In M. Botti et al. (Eds.), *Proceedings of CEPE 2009 – Eighth International Computer Ethics and Philosophical Enquiry Conference* (pp. 425-437). Kerkyra, Greece: Ionian University.

[16] Piaget, J. 1932. *The moral judgement of the child*. Routledge and Kegan Paul, London.

[17] Πλάτων [Platon]. 1981. *Θεαίτητος [Theaitetos]*. Ι.Ζαχαρόπουλος [I. Zacharopoulos], Αθήνα [Athens].

[18] Πλάτων [Platon]. 1992a. *Πολιτεία [The Republic]*. Κάκτος [Kaktos], Αθήνα [Athens].

[19] Πλάτων [Platon]. 1992b. *'Απολογία Σωκράτους [Apology of Socrates]*. Κάκτος [Kaktos], Αθήνα [Athens].

[20] Sunstein, C. R. 2005. Moral heuristics. *Behavioral and Brain Sciences*, 28, 531-573.

[21] Wallace, W. and Allen, C. 2009. *Moral machines: Teaching robots right from wrong*. Oxford University Press, New York.

The Issue of Moral Consideration in Robot Ethics

Anne Gerdes
Associate Professor
Department of Communication and Design
University of Southern Denmark
4565501323
Gerdes@sdu.dk

ABSTRACT
This paper discusses whether we should grant moral consideration to robots. Contemporary approaches in support of doing so centers around a relational appearance based approach, which takes departure in the fact that we already by now enter into ethical demanding relations with (even simplistic) robots *as if* they had a mind of their own. Hence, it is assumed that moral status can be viewed as socially constructed and negotiated *within* relations. However, I argue that a relational turn risks turning the *as if* into *if* at the cost of losing sight of what matters in human-human relations. Therefore, I stick to a human centered framework and introduce a moral philosophical perspective, primarily based on Kant's *Tugendlehre* and his conception of duties as well as the Formula of Humanity, which also holds a relational perspective. This enables me to discuss preliminary arguments for moral considerations of robots.

Categories and Subject Descriptors
K4 [**Computers and Society**]: Ethics.

General Terms
Design,Theory.

Keywords
Moral consideration, ethics of robotics, duties, as if.

1. INTRODUCTION
In a recent report on lethal autonomous robot systems, Heynes points to that personhood is what links moral agency to responsibility [11]. But is that necessarily the case, or is Heynes being species chauvinistic? The answer could well be a yes, since robots have started to come into our social lives and we interact with them in human-like ways, as if they had inner mental states. On this background, it seems that we have good reasons to dwell upon our concepts of moral agency and patiency. Especially since our interactions with, and reactions towards, robots also concerns our self-image. First, I discuss the possibilities of artificial moral agency and patiency and explore whether this counts in favour of anchoring the question of moral status in phenomenological observations of how we form relations with robots; the so called *relational turn*, favoured by Coeckelbergh [3] and Gunkel [9], who summarizes the idea as an alternative to standard explanations, which sets out to decide, who (or what) deserves moral standing on the basis of ascribing properties to the entity in question. Hence, according to Gunkel, the relational "*..alternative [..] approaches moral status not as an essential property of things but as something that is socially negotiated and constructed in face of others."* ([10]:13)

I sympathize with the relational turn, but still find that it is challenged by the fact that, over time, our human-human relations may be obscured by human-robot relations. Currently, it may seem reasonable to skip discussions about what a robot *really* is and instead focus on how it appears to us and how we engage with it by applying *as if* approaches. But in the long run, our experiences with robots may radically alter our *Lebenswelt*. Here, I'm in alignment with the ideas of Turkle [18], who fears that we may lose something of great importance if we turn to robots or even end up preferring robots over humans.

For that reason, I outline a Kantian moral argument in emphasizing his treatment of duties in the doctrine of virtues, *The Tugendlehre*, which is presented in the second part of *The Metaphysics of Morals* [13]. Related to Kant's analysis of duties, there is room for a relational perspective, which can be expressed via the Formula of Humanity. Moreover, I also make reference to virtue ethical reflections in general. Thereby, I am able to put forward preliminary arguments for granting degrees of moral consideration to robots without risking that we gradually lose sight of our folk intuition and lived experience with what it is to enter into social relations. As such, I prefer to stay within a human centered framework, even though I agree with the proponents of the relational turn that there are baffling problems inherent to this kind of mind-morality perspective. However, the mere fact that things are complicated and problems unsolved does not constitute a proper reason for rejecting a framework.

2. ROBOTS IN THE MORAL SPHERE
The role of robots in moral discourse has been widely debated both within science fiction, philosophy and science. Hence, The World Robot Declaration was issued in Japan in 2004 and within the last decade, humans have increasingly interacted with care bots, pet bots, robot toys and robots for various therapeutic purposes (see for instance [18], [6], [1]).

ETHICOMP, September 7-9, 2015, Leicester, United Kingdom.

One of the first to include robots in the moral sphere was Asimov, who issued his famous laws of robotics, which he used in science fiction novels to illustrate ethical dilemma situations in human robot interaction. From an engineering point of view, in *Moral Machines – Teaching Robots Right from Wrong,* Wallach and Allen [21] present the promises of machine morality from an engineering perspective by distinguishing between top-down, bottom-up and hybrid approaches to programming morality. Here, the first mentioned system suggests the implementation of formalizations of a given moral philosophical theory, whereas a bottom-up system requires neural network models, which gradually build up moral understanding by trial and error based performance optimization techniques. However, pure bottom-up systems are challenged by the lack of a guiding ethical theory, and as such there is no guarantee that a robot will develop a preferred kind of moral maturity. On the other hand, a hybrid model, which Wallach and Allen speak in favour of, combines these ideas from a virtue ethical outlook: Here, artificial moral agency might be obtained by integrating bottom-up learning scaffolded by top-down rules.

By the same token, from a philosophical angle, Verbeek [20] grasps the possibility of artificial moral agency by viewing technologies as mediating devices, which serve as morally active in shaping human understanding and action in the world. Consequently, even though technological artifacts do not hold human-like intentions, it can make sense to refer to distributed or hybrid intentionality and hence assign intentionality to technology in the sense that technological artifacts may play a directing role in our actions and experiences ([20]:57). Correspondingly, in moving beyond an anthropocentric understanding of agency, Floridi and Sanders [8] reject free will and mental states as necessary conditions for moral agency. On the contrary, they argue that moral agency may be assigned to intelligent artificial agents (AAs) to the extent that such AAs are interactive, i.e., able to react to stimuli by changing state, and capable of adaptive behavior as well as autonomous responses to the environment. What matters is whether an agent can perform good or evil actions, that is, whether its actions are morally qualifiable ([8]:371).

If we include robots in the moral sphere by assigning moral agency and responsibility to them, a next reasonable step would be to discuss if the time has come where we ought to discuss whether robots are worthy of moral consideration? Among others, Gunkel thinks the answer to that question might be a yes. In *The Machine Question – Critical Perspectives on AI; Robots, and Ethics*, Gunkel [9] argues that already by now the term "person" has been stretched out to include non-human agents, such as corporations. As such, we might benefit from including machines into the category of persons. If we do so, the question arises whether the kind of responsibilities we have towards robots would be on pair with the kind of responsibilities we have towards animals, corporations or other human beings?

A lot has been written about machine agency in trying to lay out how robots ought to treat humans. Typically interest centers on how we may protect ourselves from possible harm caused by robots. At the same time little has been said about machine patiency. ([9]:103]). Hence, according to Gunkel, a claim to moral consideration, or even rights, may arise based on our social interactions with robots. We design artificial companions with whom (or which) we do engage and bond. Our machines are no longer tools, but have instead gradually turned into social actors or social interactive objects. Consequently, it may be about time we begin to think about moral obligations towards robots, maybe even in the strong form of robot rights. The mere fact that Paro, the seal care robot, is not a consciousness being with inner mental states does not automatically justify that we should not grant moral consideration to Paro. Moreover, our ways of living with robots is not just about what we do with robots, but also concerns our self-perception – what do I become through the kind of relations I form with robots?

A contrast to the relational view can be found in the work of Sparrow [17]. He presents a so-called Turing Triage Test which allows him to illustrate that we would always chose a human life over a robot's life, regardless of how advanced the robot might be. The mere fact that we can never know what the robot is *really* feeling, and if it feels anything at all makes it implausible to talk about, for instance, 'punishing' a robot: "*Our awareness of the reality of the inner lives of other people is a function of [..] "an attitude towards a soul"".* ([17]:211). According to Sparrow, there exist an unbridgeable gap between reality and appearance ([17]:210).

On the other hand, Coeckelbergh, like Gunkel, suggests a relational turn and continues by arguing in favour of replacing "*..the question about how "real" or how "moral" non-human agents are by the question about the moral significance of appearance."(*[5]:181).

He displays problems with what he coins "a property approach to moral status assignment"*,* which seems to rest on the assumption that we can settle issues about moral significance with reference to a set of properties (e.g., mental states, speech, consciousness, intentionality). In this manner, we can supposedly establish a firm ground for separating out entities worthy of moral standing. But, Coeckelbergh points to problems inherent in this line of argument. Especially, it appears to be impossible to establish which properties we exactly need in order to be able to assign moral status to an entity. Also, the whole endeavor is challenged by "the other minds problem" - i.e.; the fact that we can never know for sure anything about the inner lives of others. Instead, Coeckelbergh focuses on our perceptions of robots and the way this affects our interactions with such entities:

"My suggestion is that we can permit ourselves to remain agnostic about what 'really' goes on 'in' there, and focus on the 'outer', the interaction, and in particular on how this interaction is co-shaped and co-constituted by how AAs [artificial agents] appear to us, humans ([5]: 188)

Coeckelbergh's phenomenological conception reflects a relational perspective, which takes departure in the observation of our mutual dependency. This fundamental precondition – with which everyone is actually familiar – forms a central point in Coeckelberg's so-called relational ontology, which assumes that "*relations are prior to the relata"*([3]:45), and thereby view robots and humans as "relational entities". For that reason, Coeckelbergh emphasizes a social-relational approach to moral consideration ([4]:219). But, here, unlike Coeckelbergh, I shall be arguing that we need not lean against appearance in combination with a social relational ontology. Instead, I point to a Kantian outset, which emphasizes how we can have duties *to* others and *with regard to* non-humans. Before moving forward, I find it important to stress that this paper does nothing else than provide a tentative outline of my preliminary ideas. In that respect, and all though I have reservations towards their positions, I find the work of Coeckelbergh and Gunkel highly inspiring and thought provoking.

3. AS IF

Appearance is closely related to the notion of 'as if', which is also explicitly noted by Coeckelbergh in mentioning that we interact with e.g., humanoid robots or artificial companions *as if* they could be trusted, blamed or loved. Therefore, Coeckelbergh calls for a phenomenological starting point in the investigation of human-robot relations, which takes departure in the "*observed or imagined"* human-robot relations ([5]:184).

It makes good sense to turn to analogical reasoning or to introduce *as if* constructions when confronted with unfamiliar territory. This kind of idealization, or way of using representations as tools, has been given a thoroughly treatment in Vaihinger's influential book *The Philosophy of as if* [19] in which he illustrates how fictions, i.e. *as if*-models and constructions may inform science and philosophy.

Fictions are applied due to their utility, meaning that they are justifiable when proving useful in practice. But, they are not on pair with hypotheses, which can be proved or verified ([19]: xlii). Obviously, there are shades of pragmatism in Vaihinger's work on the philosophy of *as if*. But we are not dealing with the pragmatic conception, which implies that what is useful to believe is true, since here "useful to believe" may involve *both* that which is true or false. In opposition to this, the guiding principle in Vaihinger's philosophy is the observation that fictions are not just false but contradictory. Hence, fictions are errors, but fruitful errors. Yet, Vaihinger warns us that the use of fictions may also lead us astray, hence in legal practice women used to be treated *as if* they wore minor, which caused grave injustice ([19]:148).

However, fictions are widely used in everyday thinking as well as in science, philosophy, economics, legal practice and in the description of abstract objects ect.. For instance, Vaihinger mentions Adam Smith's *Wealth of Nations*, which apply the fiction that human nature is driven by rational egoism. This fiction forms the foundation of Smith's theory. Likewise, Also, Kant, in his treatment of rational agency, requires us to act *as if* we were free even though this is not the case in the real, phenomenal world. By the same token, the categorical imperative demands that you "*act as if the principle of your action were, through your will, to become a general law of nature"* ([19]:292). Hence, according to Kant, our *vernunftbegriffe* are fictions since they do not refer to objects in the world of experience [14]:KrV B799). Actually, in explaining the role of *as* if, Vaihinger points to the fact that the term "heuristic fictions" was coined by Kant:

"Kant introduces a new term for what [..] he subsequently called "heuristic fictions": he calls the ideas "regulative principles of pure reason": they are not "constitutive" principles of reason, i.e. they do not give us the possibility of objective knowledge either within or outside the domain of experience, but serve "merely as rules" for understanding by indicating the path to be pursued within the domain of experience. By providing imaginary points on which it may direct its course but which can never be reached because it is outside reality." ([19]:273)

Also, Coeckelbergh notes that we can never have access to reality, mental states or the minds of others'. But, as noted above, instead of a mind morality approach, he suggests an alternative route. Rather than discussing the moral significance of either human or robot, we must turn to the study of appearance and relations in situations involving moral considerations in human-robot interactions ([4]:215). Consequently, when people, now or in a near future, start to treat humanoid robots as if they were moral agents, we could benefit from letting these observations guide our investigations by focusing on how humans experience and form interactions with robots through *as if* approaches.

Nevertheless, according to Vaihinger, fictions are only justifiable, not probable hypotheses. As such, I doubt that we need to take a full relational turn and introduce a social relational ontology. To me, it seems that the relational *as if* approach is challenged by the fact that, over time, our human-human relations may be obscured by human-robot interactions. Currently, it might seem reasonable to skip discussions about what robot*s really* are and instead focus on how they appear to us and how we engage with robots in social situations by applying *as if* approaches and ascribe human-like agency to them. But in the long run, our experiences with robots may radically alter our *Lebenswelt* and by then we will no longer be able to make use of *as if* approaches, because we have forgotten what human-like relations are, that is: we have become unable to 'measure' experiences up against the benchmark of human relations. Here, I am in alignment with the ideas of Turkle [18], who fears that we may let go of fundamental values, such as trust and friendship, if we turn to robots or even end up preferring robots over humans:

"At the robotic moment, we have to be concerned that the simplification and reduction of relationships is no longer something we complain about. It may become what we expect, even desire."([18]:295).

Likewise, if philosophers take departure in observed and imagined human-robot relations, they risk turning the *as if* into *if* ([19], [7]:9) and thereby lose sight of what originally constituted human-human relations.

4. A HUMAN CENTERED PERSPECTIVE

In *Robot Futures* [16], Nourbakhsh describes a future scenario in which some kids act with great cruelty towards a robot dog. The scenario reminiscences about children's abusive behavior towards animals, and the son in Nourbakhsh's story remarks that: *"These people…they're sick. Let's go home!"* ([16]:54). By the same token, Nourbaksh reports a more recent experience with an autonomous tour-guide robot, which people would get great fun from teasing while it was guiding guests visiting a museum. Nobody seemed to care when it said: "please step out of my way", it was not until the engineering team changed the phrase to also include the people being guided by the robot, that people's attitudes towards the robot were changed to the better - *even slow robots will be treated well by people when they are wrapped into a human social context* ([16]:58).

As discussed above, a justification of moral consideration to robots may rest upon the observation that once we start ascribing agency to robots, we may possible become ethical obliged towards them. Moreover, the way we treat robots will have an impact on our moral habitus. In order to take this into account, I choose to I introduce Kant's distinction between two kinds of duties, as duties *to* human beings and duties *with regard* to non-human beings and entities [13].

Consequently, in what follows, I shall be introducing a perspective, which of course, within a relational ontology, is viewed as flawed due to problems derived from this kind of anthropocentric line and its inherent "property approach to moral status ascription" [3]. Both Coeckelbergh and Gunkel argue that we need to move beyond the assumptions of mind morality philosophers. They in particular point to the vagueness of metaphysical concepts and the fact that there is no consensus on

what these concepts designate. Moreover, complications also arise from the fact that we do not have access to others' minds. Hence, the argument goes that we must rethink moral agency and patiency by turning to their alternative relational paradigm ([9], [3]).

But, in contrast to their approach, I think that one cannot reject the role of metaphysical concepts, such as consciousness, intentionality and freedom, with reference to the fact that complicated issues have not yet been settled. This would be like discharging logic on the basis of Gödel's incompleteness theorems.

Hence, In *Facing up the problem of consciousness* [2], Chalmers notes that consciousness is the outmost puzzling problem in the science of mind ([2]:200). He has coined the terms *the easy problem* and *the hard problem* of consciousness in referring to the fact that we already know about the part of consciousness dealing with e.g., our ability to categorize, discriminate, associate and recognize patterns. Additionally, over time, our knowledge about brain processes will gradually increase, and we will probably end up knowing all there is to know about the complexity of the brain. This is *the easy* problem. But, *the hard problem* of consciousness is the problem of experience, that is, to learn why all that processing accompanies my consciousness experience. As such, mental qualia escape reduction to biophysical matters, and in modern dualism, property dualism holds that the mind has two fundamentally different types of properties, bio-physical and qualia. According to Chalmers, despite interesting and advanced cognitive science and reductionist models "*the mystery of consciousness will not be removed.*" ([2]:221). As an alternative, Chalmers sets out to outline a nonreductive theory of consciousness, which I'll not go further into here, where I only wish to point to Chalmers' observation that : "*The hard problem is a hard problem, but there is no reason to believe that it will remain permanently unsolved"* ([2]:218).

By itself, the observation that the concepts of mind pose baffling problems is no argument for dismissing the project of mind philosophy. I argue in favour of re-instantiating the mind-morality perspective, which allows me to move on to a Kantian and virtue ethical perspective, in which there is room for arguments for moral consideration of robots as different from humans, as well as from other artifacts or tools.

Moreover, Kant's Formula of Humanity reflects a relational perspective in describing how we ought to treat others (persons) as ends in themselves, where by "ends" Kant means "*only the concept of an end that is also a duty, a concept that belongs exclusively to ethics.[..]."* ([13]: 6:389). As such, we can only have duties *to* human beings, since duties require being capable of obligation ([13]:192). Meanwhile, Kant's *Tugendlehre* [13] allows for a description of moral obligations *with regard to* other beings or entities. Actually, Kant gives similar reasons as above in emphasizing that a prevalent argument for having indirect duties *with regard to* non-human entities and animals rest upon our duties *to* ourselves:

"§17 [..] a propensity to wanton destruction of what is beautiful in inanimate nature [..] is opposed to a human being's duty to himself; for it weakens and uproots that feeling in hum, which, though not of itself moral, is still a disposition of sensibility that greatly promotes morality or at least prepares the way for it[..]. With regard to the animate but non-rational part of creation, violent and cruel treatment of animals is far more intimately opposed to a human being's duty to himself, and he has a duty to refrain from this; for it dulls this shared feelings of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other men. [..] – Even gratitude for the long service of a horse or dog belongs indirectly to a human being's duty with regard to these animals; considered as a direct duty, however, it is always only a duty of the human being to himself."([13]: 6:443)

Thus, a Kantian perspective, as formulated in his doctrine of virtues, enables us to introduce degrees of moral consideration along a continuum stretching from, e.g. simple artifacts, such as tools, over to, for instance, paintings and historical buildings. We have varying degrees of duties *with regard to* such entities: One could say, that I have a duty towards tools, such as for instance my garden kit, in the sense that I handle these objects with care, i.e.; I clean them after use, oil them when needed and so on. In that sense, the practice surrounding gardening includes taking proper care of one's tools, and if I fail to do so, I will either feel bad about myself and improve my behavior or continue acting carelessly. In that case, others might blame me for neglecting my duties as a gardener. Here, we are of course dealing with moral consideration in a minimal sense thereof. But, from a virtue ethical perspective [15], the way I succeed or fail in my role as a gardener is nevertheless important for my personal flourishing.

Likewise, but on a more serious scale: when confronted with acts of vandalism, for instance the destroying of historical buildings by Islamic State, we find that such acts are wrongful due to the lack of moral consideration to these architectural pearls.

We do not have duties *to* animals, but we have duties *with regard to* animals. This is so, primarily because animals deserve moral consideration because they can suffer and because the way we treat animals will influence our self-perception. Moreover, according to MacIntyre:

"*To acknowledge that there are [..] animal preconditions for human rationality requires us to think of the relationship of human beings to members of other intelligent species in terms of a scale or a spectrum rather than of a single line of division between 'them' and 'us'"* ([15]:55)

Again, the question arises: what do I, or we, as a moral community, become if we abuse animals? This indirect argument for moral consideration has been criticized by Coeckelbergh [4]:213) with reference to that it seems contra-intuitive to justify moral consideration by referring to our own well-being rather than to the well-being of the receiver of moral consideration. But, as illustrated above, actually both Coeckelbergh and Gunkel stresses the importance of a relational turn (social relational ontology) with reference to that living with robots will change our lives, hence we need to reflect upon what we become from interacting with robots. By the relational turn Coeckelbergh de-individualizes the concept of a person and holds that we have to be viewed as *relational entities whose identity depends on their relations with other entities* ([4]:215).

In addition Coeckelbergh problematizes the fact that virtue ethics faces the problem of application. Hence, we cannot establish, or delimit, what the virtues are, which ought to guide our lives, and we cannot point out precisely which entities we should grant moral consideration by exercising virtuous behavior towards them. This is a classic line of argument against virtue ethics, which has been countered by Hursthorse [12] in arguing that an ethical normative theory does not necessarily have to deliver the right answers as such, or, in the case of virtue ethics, provide a

complete catalogue of virtues. As such, a plausible normative ethical theory should not give us universal rules to guide our behavior. Instead, it should be sufficiently flexible to allow for different moral outcomes by taking into consideration relevant elements in a particular context. Consequently, when faced with dilemma situations in real life contexts, it might well be the case that two persons solve a dilemma differently. This is not a relativist standpoint, since it does not imply disagreement about the fact that there is a conflict of values, rather it takes into consideration that, in the given context, there might be more than one solution, which is in accordance with that, which is virtuous.

Thus, from a virtue ethical perspective, we develop to become what MacIntyre calls *independent practical reasoners* [15]:158) through our upbringing and through participation in moral communities, which stand as morally robust and sound practices because they are open to critical reflective examination by members from in and outside the given community.

Within this kind of human based social framework, it might still be possible to grant moral consideration to robots by introducing a continuum on a scale above artifacts - such as tools and things, which we handle - over to animals. Probably below living entities, like animals, we may place robots with which we do form *as if* social relations.

I too hold that living with robots will change our lives. But I doubt that we need to take the relational turn.

5. CONCLUDING REMARKS

Since, we already by now interact with humanoid robots, and even rather simplistic types of robots, as if they were moral agents; we ought to start deliberating about moral status. This observation might lend support to a relational turn, which allows for viewing robots and humans as relational entities, rather than subjects and objects, thereby assuming that morality is always already situated in the social sphere and phenomenologically rooted in mutual dependency between social actors – "*relations are prior to the things related*" ([3]:110). Moreover, we ought to pay attention to how human-robot interactions actually unfold, that is, focus on *appearance* or how we apply *as if* approaches when we enter into human-like relations with robots. Thus, if we follow suit with the relational turn, we might benefit from not having to struggle with the problems of property ascription and mind-morality. Even better: Coeckelbergh holds that he does not want to give up on folk intuition reflected in the idea that there is a special relation between humanity and morality ([5]:181).

Yet, in the long run, our experiences with robots may radically alter our *Lebenswelt*. Therefore, by taking the relational turn, I think we risk losing sight of something of great value to our humanity, perhaps without recognizing that this has been the case. Instead, I suggest staying within a human-centered framework. Here, I present a Kantian relational perspective, which distinguishes between others, *to* whom we have duties, and non-humans, such as robots, with *regard to* which we have duties.

Even though I place myself in (humble) opposition to the work of Coeckelbergh and Gunkel, I am deeply inspired by them. Compared to their thoroughly analyses in the field of ethics of robotics, my contribution represents nothing more than a preliminary note. For now, I have no fully fleshed out solution to offer regarding how to establish a continuum, which enables us to grant various degrees of moral consideration to non-humans. Nevertheless, when speaking about robots, I still find it worth being anthropocentric for the reasons given above, but also bearing in mind that morality is deeply linked with mortality.

6. ACKNOWLEDGMENTS

I am grateful to my dear colleague, Klaus Robering, for inspiring discussions about moral philosophy as well as for his suggestions, which helped me develop this paper.

7. REFERENCES

[1] Bartneck, C., Van der Hoek, M., Mubin, O., Al Mahmud, A. 2007. Daisy, Daisy, Give Me Your Answer Do! Switching off a Robot. *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction*. Washington DC. . DOI: 10.1145/1228716.1228746. 217-222.

[2] Chalmers, D.J. 1995. Facing up the Problem of Consciousness. *Journal of Consciousness Studies* (2): 3, 200-219.

[3] Coeckelbergh, M. 2012. *Growing moral relations: critique of moral status ascription*. Palgrave Macmillan, NY.

[4] Coeckelbergh, M. 2010. Robot rights? Towards a social-relational justification of moral consideration. *Ehtics Inf Technol.*12, 209-221.

[5] Coeckelbergh, M. 2009. Virtual moral agency, virtual moral responsibility: on the moral significance of the appearance, perception, and performace of artificial agents. *AI & Society*. 24, 181-189.

[6] Dautenhahn, K. 2007. Socially Intelligent Robots: Dimensions of Human-Robot Interaction. *Philosophical Transactions: Biological Sciences*, Vol. 362, No. 1480, (Apr. 29, 2007). 679-704.

[7] Fine, A. 1993. Fictionalism. *Midwest studies in philosophy,* XVIII.1-18.

[8] Floridi, L., Sanders, J. W. 2004. On the morality of artificial agents. *Minds and Machines*. 14(3), 349-379.

[9] Gunkel, D. J. 2012. *The Machine Question – Critical Perspectives on AI, Robots, and Ethics*. The MIT Press. MA.

[10] Gunkel, D. J. 2014. The Other Question: The Isssue of Robot Rights. *Proceedings of Robo-Philosophy 2014. Sociable Robots and the Future of Social Relations.* Frontiers in Artificial Intelligence and Applications. IOS Press

[11] Heynes, C. 2013. Report of the Special Rapporteur on extrajudicial summary or arbitrary executions on Lethal Autonomous Robot Systems. A/HCR/23/47 fhttp://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf.

[12] Hursthouse, R. 1999. *On Virtue Ethics*. Oxford University Press. Oxford. NY

[13] Kant, I. 1991. *The Metaphysics of Morals*, transl. by M. J. Gregor. Cambridge University Press.

[14] Kant, I. 1785. Akademieausgabe, vol. IV *Grundlegung zur Metaphysik der Sitten*. http://www.korpora.org/Kant/aa04/Inhalt4.html

[15] MacIntyre, A. 1999. *Dependent rational animals: Why human beings need the virtues.* Carus Publ. Company. Chicago.

[16] Nourbakhsh, I. R. 2013. *Robot Futures*. MIT. Cambridge. MA.

[17] Sparrow, R. 2004. The Turing Triage Test. *Ethics and Information Technology*. 6, 203-213. DOI: 10.1007/s10676-004-6491-2.

[18] Turkle, S. 2011. *Alone Together – Why We Expect More From Technology and Less From Each Other*. Basic Books, NY.

[19] Vaihinger. H. 1924. *The Philosophy of as if*. Transl. by C. K. Ogden. London.

[20] Verbeek, P. P. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. The University of Chicago Press.

[21] Wallach, W., Allen, C. 2009. *Moral Machines – Teaching Robots Right from Wrong*. New York: Oxford University Press.

.

Implementing an ethical approach to big data analytics in assistive robotics for elderly with dementia

Heike Felzmann
Centre of Bioethical Research & Analysis
NUI Galway
Galway, Ireland
Heike.felzmann@nuigalway.ie

Timur Beyan
Insight Data Analytics Institute
NUI Galway
Galway, Ireland
Timur.beyan@insight-institute.ie

Mark Ryan
Centre of Bioethical Research & Analysis
NUI Galway
Galway, Ireland
M.Ryan1@nuigalway.ie

Oya Beyan
Insight Data Analytics Institute
NUI Galway
Galway, Ireland
Oya.beyan@insight-institute.ie

ABSTRACT
In this paper, we analyse the ethical relevance of emerging informational aspects in robotics for the area of care robotics. We identify specific informational characteristics of contemporary and emerging robots, especially the fact of their increasing informational connectedness. We then outline specific ethical considerations arising in the design process in the H2020 project MARIO which aims to develop a care robot for persons with mild to moderate dementia in home and residential care settings. Ethical considerations regarding specific functionalities of the proposed care robot are outlined.

Categories and Subject Descriptors
K.4.1 [**Public Policy Issues**]: Ethics;

K.4.2 [**Social Issues**]: Assistive technologies for persons with disabilities

General Terms
Design, Human Factors.

Keywords
Care robotics, information ethics, privacy, value-sensitive design

1. INTRODUCTION
Big data analytics in the area of health care is currently considered to be one of the most promising innovative approaches to increasing knowledge of health factors and, ultimately, to improving the delivery of health care. Health care stakeholders now have unprecedented types and quantities of data at their fingertips. The area of care for the elderly is one of the areas where big data analytics might contribute substantial improvements; ICT solutions like assistive robotics and ambient assisted living (AAL) for the elderly hold the promise of supporting independent living for the elderly beyond the stage at which currently more intense forms of monitoring and care, often in quite restrictive residential settings, is considered necessary. However, while big data has a huge potential to create significant value, it also contributes to qualitatively new concerns with regard to the use of personal information.

In this paper we will present considerations in addressing information related ethical issues in the development of a particular assistive care robot within the European H2020 project MARIO ("Managing active and healthy aging with use of caring service robots"). The project aims to develop an assistive care robot for persons with mild and early moderate dementia. These service robots will be used to support users in retaining their health and ability to perform activities of daily life, and increase their social connectedness and resilience, thereby mitigating the effects of dementia. The goal is to allow persons with dementia to stay living independently in the community for as long as possible.

2. CARE ROBOTICS: AN ETHICALLY SENSITIVE FIELD
Care robotics is a field of robotics that has been emerging over the last decade as a response to demographic developments in the developed world. Countries like Japan have pursued the use of robots in elderly care for a long time. Europe is now pursuing similar developments, with the European research agenda including care robotics for the elderly as a part of their strategies for aging, and the European Strategic Research Agenda for Robotics in Europe 2014-2020 (SPARC) identifying assisted living robots as part of the growing market of consumer robots [13]. Similarly, the UK Robotics and Autonomous Systems (RAS) strategy RAS 2020 includes reference to health and social care robotics for the elderly population [11].

Despite its strategic endorsement as promising area of technological innovation, there has been significant unease with the introduction of care robotics into elderly care settings. Most prominently, and frequently mentioned in strategic documents, concerns centre around the changes the introduction of robots bring to the nature of care, in particular the potential dehumanisation of care and the replacement of caring interpersonal relationships with machines. Most documents acknowledge that these concerns need to be addressed with sensitivity for care robotics to gain social acceptability.

Ethically speaking, these concerns are ultimately about the question whether core values of care can be realised when care robotics enters the picture, and if so, under what circumstances ().

ETHICOMP2015, September 7–9, 2015, De Montfort University, Leicester, UK.

The nature of the relationship between robots and humans is at the centre of these concerns. In their influential review article Sharkey & Sharkey [12] have set out six core problems in relation to elderly care robotics: (1) the potential reduction of the amount of human contact the elderly person receives, as care is being delegated to the robot; (2) the potential increase in feelings of objectification and loss of control of the elderly person due to robot monitoring and standardised intervention into their activities of daily life without being part of a mutual relationship in which the relationship can be re-defined and re-negotiated; (3) a loss of privacy, due to continuous monitoring and recording in their daily life of their activities and expressions by the robot; (4) a loss of personal liberty due to restrictive interventions by the robot; (5) deception and infantilisation due to uses of robots that may foster the build-up of attitudes that are not appropriate to the robot's actual characteristics and capabilities (e.g. beliefs regarding the robot's emotional relationship to and care for the person) or may not do justice to their human dignity (e.g. through provision of interactive opportunities or physical features appropriate for small children rather than adults); (6) the question with regard to the circumstances in which elderly people should be allowed to control robots whose purpose may include monitoring, and behavioural interventions like reminding, activating, incentivising which for optimal effect would require functioning independently of the elderly person's mood and preferences.

What is evident from this list, as well as many other writings on the ethics of care robotics, is that while issues relating to the ethics of information are addressed and certainly implicitly present, they are significantly less prominent than the aspects of personal dignity and the nature of the relationship between robot and elderly person. The focus on the human-robot relationship is not surprising given that ethical care is generally described as an essentially interpersonal phenomenon. We do not intend to question the fundamental significance of realising ethical care, but what we aim to do in the following is to further foreground the informational aspect, especially in light of recent developments in the field of ICT that have transformed, and are continuing to transform, the informational functions of robots..

3. THE USE OF INFORMATION IN CARE ROBOTICS

In robot ethics, ethical issues relating to the use of information have been less strongly emphasised in the literature. However, as we argue here, the informational dimension is becoming increasingly more complex and significant, as robots in general, but especially most care robots have transformed from tools designed for highly specific, often physical tasks to multipurpose information hubs that are highly connected with their environment and have highly complex distributed information flows as essential characteristic of their functioning.

One obvious core concern with regard to the use of information is the issue of informational privacy. Privacy concerns are much discussed in the literature on all information technologies, and have been for some time. Their significance is evident also in the widespread awareness among laypersons of privacy as an important issue in the field of ICT. In order to appreciate the particular meaning of privacy concerns in the context of care robotics, it is essential to understand recent technological and functional developments for care robots and consider questions of privacy in the particular context of the robot's functioning in everyday life on the basis of its informational architecture. As Nissenbaum [8, 9] has elaborated in her influential contributions to the debate, privacy needs to be considered in relation to the specific contexts of use, where information practices and privacy expectations may differ significantly. In the following we will first outline the complexity of the informational architecture in current care robots and then discuss how these considerations manifest themselves in relation to specific care robot functionalities envisaged in the MARIO project.

3.1. Robots as information hubs

Alaiad and Zhou define privacy concerns as the stakeholders' lack of control over the collection and use of their personal information after they have adopted the system [1]. Despite the fact that robots are perceived as independent entities by their users, they generally communicate with many other systems. Many care robot functionalities may require storage of data and the comparison of data with other systems such as electronic health records, or they may repurpose the stored data to improve their intelligence. With new advances in pervasive computing and ambient assisted living environments, the sharing of personal data between robots and components of smart environments has already begun and will increasingly become more common and pervasive. Robots as part of such an ecosystem will go significantly beyond their traditional role as stand-alone entities that facilitate specified parts of care, but instead become nodes of complex information sharing networks. Robots that are equipped with sensing and communication capabilities will interact with a wide range of sensors and distributed data sources. Core care robot functionalities, such as monitoring users to detect potential health risks, require communicating with wireless physiological sensors and accessing users' health records. Especially care robots with the purpose of facilitating independent living will increasingly interact with smart devices such as refrigerators, entertainment sets, heating systems, and become part of this pervasive informational environment, making IoT-aided robotics applications a tangible reality of our near future [5]. Companion robots similarly will increasingly use a variety of sensors and internet-based information for inferring context sensitive responses. Grieco et al. [5] highlight specifically two new advances in robot technology that will significantly change the way robots operate: IoT aided robotics applications and cloud robotics. These are increasingly redefining the robot's function and existence in distributed and pervasive environments.

IoT aided robotics applications are a digital ecosystem where humans, robots, and IoT nodes interact on a cooperative basis. The concept of the 'Internet of Things' (IoT) refers to the pervasive presence of a variety of things or objects – such as Radio-Frequency IDentification (RFID) tags, sensors, actuators, mobile phones, which are able to interact with each other and cooperate with their neighbours to reach common goals [4]. It is expected that sensor networks will become increasingly integral to the human environment, in which communication and information systems will be invisibly embedded [2]. That means that entities such as smart objects, sensors, servers, and network devices complement the robot, so that the robot and various IoT devices connect through a complex and heterogeneous network infrastructure. The robot interacts with the IoT, databases, and the internet and becomes a central node in this information network where all nodes are linked. The robot continuously interacts with the environment that is equipped with a wide range of intelligent devices and exploits this dense IoT network to fulfil tasks in a manner sensitive to changes in the environment [5].

In relation to challenges regarding the use of information, these fully decentralized and spatially distributed components raise

unprecedented challenges for data security. The distributed IoT network is more vulnerable to attacks, especially due to frequently insufficient security features of smart devices. In this architecture attackers could hijack unsecured network devices, like sensors, routers and robots, converting them into bots to attack third parties or could target communication channels and extract data from the information flow [2]. Eavesdropping over the IoT network thus becomes possible, especially as attackers could target communication channels to extract information and data. This may lead to unauthorized access to massive amounts of private information. A particular threat could be denial of service attacks that overload networks. In some care contexts even temporary unavailability of the robot, in the case of such attacks, may cause harm to users or put them in danger. Moreover specific nodes of the IoT networks, such as the robots themselves, might be captured. If robots are hacked they might pose a danger to humans and their environment. These vulnerabilities may lead to serious user safety issues as well as privacy and security concerns for data stored on or transmitted through the system.

Cloud robotics is a new paradigm in robotics, where robots can take advantage of the Internet as a resource for massive parallel computation and real-time sharing of knowledge and big data sets [7]. The cloud robotic architecture leverages the combination of an ad-hoc cloud formed by machine-to-machine (M2M) communications among participating robots, and an infrastructure cloud where resources are dynamically allocated from a shared resource pool in the ubiquitous cloud, to support task offloading and information sharing in robotic applications [6]. Recently, cloud robotics applications have begun to explore novel approaches such as creating a web community by robots for robots to autonomously share descriptions of tasks they have learned [15]. For example, the DAvinCi project used cloud computing for service robots in large environments and parallelized some of the robotics algorithms as Map/Reduce tasks in Hadoop [3].

Although cloud robotics allows robots to benefit from the powerful computational, storage, and communications resources of cloud computing, it also brings challenges regarding security and privacy. De Oliveira [10] lists three relevant kinds of risks: (i) dependency on the connection and availability of the cloud computing resources, (ii) lack of control since a number of procedures are no more under the user's control, for example backups, where it is unclear to the user by whom is it performed and where the data is stored, (iii) vendor lock-in with the consequence that migration to other products and data portability may become impossible. In the healthcare setting these disadvantages may raise serious consequences related to the safety of user, when internet connections might be disrupted, the privacy of sensitive data, and the continuity of care if providers are changed. Threats have been considered particularly in relation to data privacy and security. Cloud computing means that all the computer hardware and software used by a Cloud client (a company, a public administration or an individual) is provided by another company (the Cloud service provider, CSP) and is accessed over the internet [10]. In cloud computing systems data is stored in multiple locations by various service providers [16]. This may lead to loss of control over data and consequently results in privacy concerns. Relevant threats include disclosure by cloud computing providers of personal or confidential data to third parties, including potentially clients' competitors for monetary reasons, or the replication and use of sensitive information for data mining purposes, or the use of personal data for a variety of purposes not authorised by the data subject or lacking a proper legal basis [10].

3.2. Robots as delegated agents of the user

In this densely connected ecosystem, robots are becoming increasingly autonomous decision makers. IoT devices and sensors continuously send information to the robot, and the robot is granted the autonomy to interact with these systems and make decisions about regulating the surrounding environment. The autonomy of robots has increased not just in relation to physical maneuverability, but also in relation to increasingly complex decision-making. Users of care robots, especially persons with dementia, will increasingly delegate many aspects of their decision making to the robot. Robots dynamically interact with complex systems and with increasing functional abilities will make a wide range of decisions and apply them on behalf of the user, potentially bypassing active input by the user entirely. This raises significant issues regarding the role and use of information that underlies such decision-making. Increasingly robots do not just impact on the the physical environment of their users or provide limited, task specific information, but control the informational environment for humans more comprehensively. In the literature, ethical and privacy considerations are mainly focused on the human–robot interaction and ethical characteristics of their relationship, however, the increasingly complex role of the robot in mediating the user's informational environment raises additional concerns.

A significant ethical consideration in this context is how the robot's autonomous actions will impact on the agency of the person who it serves. In addition, the autonomous decisions of the robot impact not only on the person they are serving, but also on the surrounding environment, including other systems and persons that are related to the user. While the robot may have significant information available on preferences and needs of the primary user, this might not be the case for other persons affected by robot actions. In this context, the wider question of what requirements an autonomous robot needs to fulfil to not impact unduly on other persons. The scope of robots' autonomous intervention also needs to be carefully defined, considering at what point and to what extent the robot should autonomously perform actions on behalf of the user, or make adjustments, presumably to improve their choices. In this context, it is essential to reflect on the significance of preserving agency and autonomy for the user. Actions that might be beneficial from a health point of view might not be beneficial from the agency or dignity point of view.

One rather mundane use case in this context would be the robots' creation of healthy shopping lists, making sure a choice of healthy food items are available to the user, for example through online ordering, and making suggestions on meals and snacks for the user on the basis of the items available. This is certainly in the service of health. However, it would need to be considered how important decisions on food are for the user, whether there are certain kinds of food that have a specific significance for the user, or whether retaining the autonomy of shopping is significant for the user's self-understanding.

3.3. Robots as providers of continuous representations of the user's life

Care robots, as part of their functionality acquire comprehensive information about their owners, their immediate environment and lifestyle. The layout of the house, habits such as sleeping, exercising, third persons entering the house, appointments or

communications online are continuously recorded. Even though it is an ethical convention not to store information regarding intimate situation such as bathing, the robot still needs to store substantial amounts of sensitive personal information about daily habits such as daily routine activities, eating habits, or social interactions, in order to ensure the desired functionality. Care robots also may interact with other IoT devices for grocery shopping, securing the house and measuring user's physiological parameters such as respiratory rate or blood pressure. They also may connect with medical records which provide additional personal information about the user. All this wider, distributed landscape of data is becoming integrated into the robot intelligence and provides more complex and comprehensive information than the robot itself could capture with its built-in monitoring devices.

In care settings that support users with dementia, robots are also loaded with data that may be used to address the memory impairments. Robots may store memories of significant people in the user's life, especially in the form of photos or videos, store information on their interests, such as the music they loved, the sports team they are following, or other hobbies and passions that they have been pursuing during their lives.

Moreover users need to be reminded about the people they know. To fulfill this requirement robots gather not only information about the user they are serving, but also about their families, neighbors, and friends. The names, faces, addresses and additional information about their relationship to the user is stored in the memory of the robot. With the robot's internet connection they can also be followed with social media, to supplement the stored information.

When robots collect all this information that is directly related to the private sphere of the lives of its users, it may accidentally or intentionally disclose such information to a third party. Syrdal, et al. studied robot users' feelings and concerns in case of an accidental information disclosure with the service robot PeopleBot [14]. The study showed that most of the participants felt uncomfortable about the robot sharing personal information in social settings without having control over such disclosures. Participants considered information about their personality and other psychological characteristics as sensitive. They also raised the concern about someone else's robot collecting information about them and using it. As robots become a part of smart living environments, they are further extending their observation capacity by communicating with other devices and by autonomously searching the internet, and they will collect much more in depth sensitive information not only on their owners, but also others who have connections to their owners and/or may be in the range of the robot's recording capacities. Neighbours living next door, an old friend from photos, family members will be entered into the robot's storage, frequently without their knowledge or agreement. Although all this information may serve a valuable purpose, the aggregation of significant amounts of such information is intrinsically problematic. Disclosure of such information, whether intentional or accidental, is only one issue. Due to its connected and searchable nature such information storage is significantly different in kind from photo albums, diaries, address books or collections of memorabilia, where other persons' information may similarly be stored without their knowledge, but would not be available for further use or data mining. These potential further uses of information and the preservation of privacy need to be taken into account in the design of the informational management of the robot.

4. MARIO FUNCTIONALITIES, ETHICAL CHALLENGES, AND POSSIBLE SOLUTIONS

The MARIO project aims to develop a multifunctional care robot that will support elderly persons with mild to early moderate dementia in maintaining their independence and social connectedness. It is targeted at both home and residential care settings. It will have a range of functionalities in the area of (i) health assessment and monitoring, (ii) reminders and instructions for activities of daily life, (iii) entertainment and hobbies, (iv) reminiscence and social contact. User preferences regarding these functionalities have been elicited from persons with dementia as well as formal and informal carers in three trial sites (in residential care settings in Galway, Ireland; in a geriatric unit at the IRCCS hospital Casa Sollievo della Sofferenza in San Giovanni Rotondo, Southern Italy; and in a community setting in Stockport, UK). However, the precise definitions of functionalities to be included has not been finalised at the time of writing, so the following considerations are still indicative. These different functionalities involve different challenges regarding the ethics of information and privacy management in particular. In the following the particular challenges and some suggested solutions for each of these categories will be discussed.

4.1. Health assessment and monitoring

Desiderata for the robot include the performance of monitoring of different health aspects, including potentially vital signs like blood pressure and some aspects of geriatric assessment. In addition, possibilities of monitoring the intake of medication and fluids, two major causes of adverse health impact in the geriatric setting, is also under discussion, although the precise technical implementation of those suggestions still needs to be determined. Such information will be transferred to the health records of the person with dementia, allowing for potentially more comprehensive and regular assessment than feasible otherwise, which would be especially desirable for persons living in the community as opposed to residential settings. One particular benefit of such regular information collection by the robot would be that health professionals assigned to the care of the user can be made aware of changes in a timely fashion, so that emerging risk factors indicating potential deterioration could be identified before adverse events take place. It is also intended to monitor and record when adverse events like falls occur. Proposals for robot functionalities in this context include the development of risk indices on the basis of such information.

However, the information processed in such assessment and monitoring activities is highly sensitive and raises data protection and privacy issues. Unlike the transfer and storage of medical data within protected internal networks for medical records in health organisations, in this case information will be transferred wirelessly and is likely to be stored in the cloud which might allow for potential data breaches at different points, especially if the robot is used in a home setting. In addition, monitoring for several of the functions will rely on video analysis and requires the processing and storing of significant amounts of rich behavioral information that is also highly identifiable. According to good data protection practice, it will need to be ensured that data recorded and especially data stored longer term is not excessive and that data processing and storage options either minimise data usage or have significant advantages over less data intensive alternatives. This principled reduction of data storage is also a core tenet of privacy by design. In particular, with a robot

that will accompany the persons throughout the day, including times of intimate activities, it will be necessary to ensure privacy and make it possible that the robot will stop recording information, especially video footage, without need for active requests by the user.

In addition, the person with dementia will need to agree to any health assessment and monitoring function before the introduction of the robot, just like informed consent is usually required for any health intervention. This consent should not be an all-or-nothing consent, but a certain degree of flexibility should be possible, i.e. users should have the option of excluding at least certain functions. (In the context of the trials, this may need to be handled more rigidly due to the importance of maximising user data on all functionalities. However, gaining some experience with a consent process that is sensitive to the users' needs is desirable for the user trials.) The initial consent will need to be facilitated by a person with competence in working with persons with dementia, as particular challenges in relation to memory and confusion might arise in the information process. Informed consent is a challenging process under any circumstance, but even more so for persons who suffer from memory problems.

More generally, the initial consent should be embedded in an adaptable dynamic consent framework where different options, including the potential switching off of monitoring functions, should be made available to persons with dementia. It is desirable that a range of carefully designed modified settings would be made available depending on the user's level of capacity and health state. In the interest of users' autonomy and privacy, it is desirable for persons to have the option of switching off some of Mario's functions (or switching off the robot altogether), at least temporarily, unless doing so would bring significant risks. For example, users who have no history of falls or severe disorientation might have the option to switch off safety monitoring functions at the very least temporarily. In this context it is essential not to assume automatically that safety and health benefits are always the overriding values; the significance of a person's dignity and autonomy may mean that at times risks are taken to realise those other values.

4.2. Reminders and instructions for activities of daily life and safety

Intended functionalities for the robot include a variety of reminders and instructions for different activities that the elderly person might have problems remembering or executing correctly. This includes for example reminders for activities that should be performed regularly, for example to take medication, take fluids, engage in physical activity, or go to the toilet. It may include reminders of scheduled activities, visits or appointments, based on calendar information. It may also include reminders (based on local weather apps) regarding appropriate types of clothing when the person is leaving the house, or reminders (based on RFID signals) of relevant items to bring, like wallet, purse or keys, and identifying where those are located. It could include reminders of the date and time of day, especially when users wake up, as they are particularly prone to being disoriented at those time. For users who wake up at night and start wandering they should also be reminded to go back to bed and/or not to leave the house, when appropriate. Users might also be reminded to switch off the hob or adjust the heating if sensor data indicates that this is required, and shopping needs may be identified based on the fridge content and recorded food preferences.

Instructions for activities of daily life include, for example, instructions on the choice of clothes and/or the sequence of getting dressed, on the choice of cutlery, on how to find the way around in the house or institution, for example how to get to the toilet, the sitting room or common room, or back to one's bedroom (some of these could also be integrated with a social activity calendar). They may also be adapted to specific needs, depending on the particular challenges that an individual encounters in their environment.

Some reminders, in particular, can be set at fixed intervals, like in a calendar, without taking into account a person's actual activities. However, to be more sensitive to the activities of the user and any situation-specific need for help, some reminders and instructions will be implemented on the basis of actual user behaviour. Like the above case of health assessment and monitoring, this includes more extensive monitoring and recording of user behaviour, which raises privacy issues, especially urgently if intimate behaviour is involved. Accordingly, it would need to be assessed whether robot functionalities are likely to involve the use of sensitive information, and for those in particular it would need to be balanced carefully whether the additional benefit of flexibility and adaptation to user needs is sufficient to offset the risk of privacy infringements.

Instructions and reminders also come with the particular challenge not just of data privacy but also of social privacy, in the sense of reminders or instructions being audible or visible to third parties in the user's social environment. Reminders on activities that users and/or others generally expect adults to be able to perform themselves, and especially reminders for intimate activities, might be considered socially problematic by the user or their social environment and have the potential for significant embarrassment. This might be particularly significant for users who are experiencing uneven loss of abilities. They may be highly functioning in many domains and have a high level of self-awareness and social integration, but may have significant difficulties with particular activities which they would prefer to keep confidential from others. Accordingly, the design of the functionalities will need to take into account this potential for embarrassment and carefully design robot intervention so that less socially intrusive modalities for reminders and instructions are used when other persons are present to maximise privacy and dignity. This may be particularly significant for residential care settings where users may not have a single room to themselves, and accordingly robot interventions may be likely to be overheard/visible to others with regard to nearly all of the user's activities and not just restricted to defined social settings.

4.3. Entertainment and hobbies

Functionalities regarding entertainment and hobbies are particularly significant for persons with dementia, as they provide opportunities for activation, positive experiences and potentially also social integration. These are all relevant for a better experience of quality of life and constitute protective factors against the further decline of dementia symptoms. Envisaged MARIO functionalities include the provision of broadcast media access, games, and social media connectivity to relevant sports clubs or other information that the user is interested in or passionate about. It may also involve assisting users in searching for further information on matters of interest. Information on hobbies and interests, like previously discussed types of information stored through the robot, is similarly personal information where care should be taken to avoid breaches of privacy.

Facilitating continued engagement with subject matters that a person has been passionate about is particular valuable for persons with dementia who may have begun to withdraw from their ordinary activities and may be entering a vicious circle of mutually reinforcing withdrawal and further loss of functioning. For certain activities, one of the potential advantages of engaging with subject matters or games via the care robot can be that loss of functioning may make playing certain games with other persons difficult, while a robot might facilitate reminders and help to the person where needed, making possible the enjoyment of activities that may otherwise not be possible for the persons. For other activities, robot mediated activity might have a social dimension in that the robot might offer activities that are not just targeted to the primary user, but may involve peers. This might include singing, listening to music, watching movies, light exercise, or engaging in games. Given the importance of increasing social connectedness in persons with dementia, it is desirable to further explore the integration of a social dimension of activities provided by care robots.

The main information related concerns in using these functionalities correspond to previous considerations in relation to other functionalities. The robot might be collecting monitoring information of others that neither are aware of that fact nor have agreed to it, and such collection needs to be minimised. The robot also needs to be able to adapt its interactions with the user to the context, especially the distinction between individual and social settings. Interactions like reminders which might be appropriate in a one-on-one setting may be embarrassing or possibly convey too much personal information in social settings.

4.4. Reminiscence and social contact

Memory impairments are a core symptom of dementia. Reminiscence activities have been shown to be particularly beneficial for persons with dementia. These involve actively engaging persons with their memories of the past, for example persons, events and locations that were important to them. It allows them to reconnect and engage with important parts of their lives, counteracting confusion and a sense of loss that may be experienced in the engagement with the present where memory impairments often have their most significant impact. In contrast, persons with dementia can generally access long term memories much more easily than more recent events. Care robots can assist or facilitate reminiscence activities, either as an aid for interpersonal engagement with a carer or family member, or as an independent, fully robot-facilitated activity. In order to fulfil such a function, a significant amount of personal information needs to be stored in the robot, including basic information on family members and crucial events, photos, videos, and family stories. This raises a number of ethical issues around the use of information. First of all, it raises the issue whether consent is needed from others to store information that connects them to the person with dementia. For reasons of practicability but also the comparability to the use of social information and photos in other private and social contexts, such consent requirements should not be too onerous. The mere fact that photos of a person are stored should not be sufficient for demanding consent; however if extensive, sensitive or personally identifying information is being stored, seeking permission for this use would likely be appropriate. (What exactly a consent requirement entails could also be dependent on factors such as what technological possibilities of uses of such information are, whether these are implemented in any way in the robot, or what this person's sharing practices on social media are.) Seeking consent for such use of information might be raising privacy issues, insofar as it implicitly requires informing the person from whom consent is sought about the extent of the memory problem that the user is experiencing. How exactly consent should be sought and who should be in charge of addressing the issue are other open questions. It might be too onerous for the person with dementia to be responsible for the process; on the other hand it might be perceived as inappropriate or potentially patronising if another person is addressing the issue for the person with dementia. In addition, this raises the issue of data security and the potential for data breaches, which in this case affects not just the person with dementia, but also those persons whose information is being stored.

In addition to reminiscence activities, such personal information about other persons will also be used for a range of functions related to social connectedness. Functionalities in this area include the connection with social media and photo sharing services, the use of Skype or similar services, or the use of face recognition software to help the person with dementia identify persons upon meeting them. All these functions rely on storage of some information about other persons, but may also involve further collection of such information, such as social media contributions or current photos. Collection of such information by the robot needs to be designed to minimise stored data, for example through explicit requirements of selection of favourite photos, rather than wholesale storage of incoming information. One further consideration in relation to the use of social media is also that such use will also be analysed by social media providers, raising further privacy issues. Connection through a specific kind of robotic device might be identifiable; characteristics of dementia might also be inferred from contributions by those providers, like other psychological characteristics. This might not only have an impact on how advertising is targeted to the person on social media, but could potentially even have wider privacy implications if the person thus becomes identifiable as a person with dementia to the provider, or even additional parties if such information is being sought and sold on by providers or data brokers.

Finally, the use of stored information for social purposes like identification of a person on the basis of face recognition or the use of reminders on the personal connection to or shared experience with the person with dementia has the potential to be highly beneficial by improving social connectedness and avoiding awkward or hurtful moments of lack of recognition of a loved one or friend. However, depending on how reminders are presented to the person, they might be noticeable to the other person or even potentially socially disruptive. Care needs to be taken in the design process to implement such reminders in a discreet or socially acceptable manner.

5. CONCLUSION

As outlined in this paper, the informational challenges arising in care robotics are substantial and increasingly relevant. The potential of adapting and further refining care robot functionalities on the basis of massive amounts of complex connected information is considerable, but informational processes need to be adjusted on the basis of careful consideration of the ethical implications of such uses of information. Maximising privacy, both in the sense of data protection and social privacy, is a core concern. Allowing users and others affected by the collection and use of personal information sufficient transparency and control is a further challenge that needs to be met. Doing justice to these informational considerations is one important precondition for achieving ethical acceptability of care robots.

6. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Horizons 2020 – the Framework Programme for Research and Innovation (2014-2020) under grant agreement 643808 Project MARIO 'Managing active and healthy aging with use of caring service robots'.

7. REFERENCES

[1] Alaiad, A., and Zhou, L. 2014. The determinants of home healthcare robots adoption: An empirical investigation. *International Journal of Medical Informatics*, 83, 11 (2014), 825-840.

[2] Alsaadi, E., and Tubaishat, A. 2015. Internet of Things: Features, Challenges, and Vulnerabilities, International Journal of Advanced Computer Science and Information Technology, 4, 1 (2015), 1-13.

[3] Arumugam, R., Enti, V.R., Bingbing, L., Xiaojun,W., Baskaran, K., Kong, F.F, Meng, K.D. and Kit, G.W. 2010. DAvinCi: A cloud computing framework for service robots. *Robotics and Automation (ICRA), 2010 IEEE International Conference*, 3084-3089.

[4] Atzori, L., Iera, A., and Morabito, G. 2010. The internet of things: A survey, *Computer networks*, 54, 15 (2010), 2787-2805.

[5] Grieco, L., Rizzo, A., Colucci, S., Sicari, S., Piro, G., Di Paola, D. and Boggia, G. 2014. IoT-aided robotics applications: Technological implications, target domains and open issues. *Computer Communications* 54, 32-47.

[6] Hu, G., Tay, W.P., and Wen, Y. 2012. Cloud robotics: architecture, challenges and applications. *Network, IEEE* 26, 3 (2012), 21-28.

[7] Kuner, J. 2010. Cloud-enabled robots. *IEEE-RAS International Conference on Humanoid Robotics*.

[8] Nissenbaum, H. 2004. Privacy as contextual integrity. *Washington Law Review* 79, 1.

[9] Nissenbaum, H. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

[10] Oliveira, P.C. 2014. *Protection of Personal Data in the era of Cloud Computing, The Internet of Things and Big Data* , 2014 . http://www.rlpdp.com/wp-content/uploads/2014/10/Paper_Data-Protection_CC_IoT_BData_final-1.pdf

[11] Robotics and Autonomous Systems (RAS) Special Interest Group 2014. *RAS 2020: Robotics and Autonomous Systems. A national strategy to capture value in a cross-sector UK RAS innovation pipeline through co-ordinated development of assets, challenges, clusters and skills*; https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20Strategy?version=1.0

[12] Sharkey, A., and Sharkey, N. 2012. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology* 14, 1, 27-40.

[13] SPARC 2014. *Strategic Research Agenda for Robotics in Europe 2014-2020*, https://connect.innovateuk.org/documents/2903012/16074728/RAS%20UK%20Strategy?version=1.0

[14] Syrdal, D., Walters, M., Otero, N., Koay, K,. and Dautenhahn, K. (2007). He knows when you are sleeping - privacy and the personal robot companion. *Proceedings of the workshop on Human Implications of Human–Robot Interaction, Association for the Advancement of Artificial Intelligence (AAAI'07)*, 28-33.

[15] Tenorth, M., Perzylo, A.C., Lafrenz, R., and Beetz, M. 2012. The roboearth language: Representing and exchanging knowledge about actions, objects, and environments. *Robotics and Automation (ICRA), 2012 IEEE International Conference*, 1284-1289.

[16] Zhou, M., Zhang, R., Xie, W., Qian, W., and Zhou, A. 2010. Security and privacy in cloud computing: A survey. *Semantics Knowledge and Grid (SKG), 2010 Sixth International Conference,* 105-112.

The invisible robots of global finance: Making visible machines, people, and places

Mark Coeckelbergh
De Montfort University
Gateway House, The Gateway
Leicester LE1 9BH, United Kingdom
Telephone +44 116 257 7487
mark.coeckelbergh@dmu.ac.uk

ABSTRACT
One of the barriers for doing ethics of technology in the domain of finance is that financial technologies usually remain invisible. These hidden and unseen devices, machines, and infrastructures have to be revealed. This paper shows how the "robots" of finance, which function as distance technologies, are not only themselves invisible, but also hide people and places, which is ethically and politically problematic. Furthermore, "the market" appears as a ghostly artificial agent, again rendering humans invisible and making it difficult to ascribe responsibility. Epistemic invisibilities thus become moral invisibilities. Finally, if we want to render finance more socially and ethically responsible, we also have to reveal the hidden efforts of many individuals and communities to re-invent finance by means of alternative financial practices and technologies. Research on responsible innovation should also consider less visible innovation that happens outside academia and industry.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Ethics

General Terms
Human Factors

Keywords
Ethics of finance; ethics of financial technologies; ethics of robotics; ethics of AI; social studies of finance; distance; invisibility; phenomenology of finance; philosophy of finance; high-frequency trading; algorithmic trading; electronic trading

1. INTRODUCTION
One of the barriers for doing ethics of technology in the domain of finance is that financial technologies usually remain invisible. Finance is generally not understood as a technological (and social) practice. It is supposed to be about people, for instance about the rights and duties of market participants [1]. Usually technologies are only attended to when they malfunction. To use Heidegger's terminology: usually they are ready-to-hand; we just use them and do not focus on the tools we use. Only when they break down technologies become present-at-hand. [2] For instance, high-frequency trading by means of algorithms only becomes visible when something goes wrong, as was the case for instance in 2012 when a computer glitch caused a company to lose millions after it started using new software [3]. Thus, usually financial technologies remain out of sight.

In order to evaluate financial technologies, therefore, the unseen machines and devices of finance have to be revealed; otherwise we remain blind to the technological developments in finance and cannot evaluate their social and ethical consequences. For this purpose we need to re-write the history of finance as including a history of financial technologies and – with the help of social studies of finance, including ethnographical work – we need to reveal the artefacts, devices, machines, and infrastructures that make possible global finance.

Elaborating arguments drawn from my recent book *Money Machines* [4] this paper reveals the "robots" of finance and argues that they are not only themselves invisible, but also contribute to "distancing" processes which hide people and places affected. I will argue that this is a problem for the ascription and exercise of responsibility. I further discuss other ethically problematic invisibilities in finance: invisible humans under the spell of the ghostly artificial agent "the market" which therefore escape responsibility (ascription), and hidden efforts of individuals and communities to develop alternative financial practices and technologies – innovations which usually do not show up in research on ethics and responsible innovation in finance.

2. INVISIBLE FINANCIAL TECHNOLOGIES
Most discussions in ethics of finance focus on human behavior and character. This is understandable. "Ethics" is usually seen as concerned with what humans do and are. But this is misleading, since such an omission hides the many ways humans and technologies interact and entangle in practices, experience, (hi)stories, and knowledge production. For example, the history of finance is usually written in a way that conceals financial technologies: it is about the history of people and financial institutions. And financial technologies do not even appear on the radar of contemporary ethics of finance.

To start changing this I have written a very brief history and phenomenology of financial technologies [4] which highlights financial technologies in the history of finance and their role in

shaping society and human subjectivity: it pays attention to such things as clay tablets, writing, money, and other financial technologies in ancient civilizations, but also to contemporary financial technologies such as electronic currencies (consider for instance Bitcoin) and trade algorithms used in so-called algorithmic trading. In addition, we can learn from social studies of finance and STS: they have been extremely helpful in revealing the social and material side of finance. For instance, Callon has studied the prostheses and assemblages of finance [5], Knorr Cetina has described the technologies on trading floors such as computer screens [6], and Beunza and Stark [7] have revealed networks of tools such as computer programs, screens, robots, cable connections behind algorithmic trade. It turns out that finance is not only about people and very abstract institutions, but also about very concrete materialities which shape finance as a social-material and technological-material practice.

Some of these technologies are relatively new, and have received comparatively little attention in recent thinking about technology. Like in other fields, contemporary electronic ICTs are used to automate trade. Let me say more about this automation in finance.

3. DISTANCING AND THE HIDING OF PEOPLE AND PLACES

Automation has significantly changed finance. Today algorithms take over many trade actions: so-called "algos" execute trading for investment banks, pension funds, and so on. In high-frequency trading algorithms are used to trade large volumes of securities at very high speed. These rather invisible "robots" contribute to what I call a process of "distancing" [4]. Since the transactions are conducted by algorithms the human has less control – there is a distance between human decision and the transaction – and in electronic environments, people, goods, and places influenced by these transactions are hidden. Today the trader works in a kind of "cockpit", very similar in kind to an airline cockpit [7]: a highly technologically mediated environment in which contact with the "reality out there" is filtered through the electronic interface. There are numbers on a screen; people have disappeared from view.

This is problematic from an ethical point of view, since it becomes difficult to ascribe and exercise responsibility. Since Aristotle [8] knowledge and control are conditions for responsibility: in order for you to act responsibly, you have to know what you are doing and you have control over what you are doing. But meeting these conditions becomes increasingly difficult when trading is delegated to algorithms and when the electronic technologies used in finance are screening off the socially and ethically relevant consequences of trading acts and decisions. For instance, if because of particular electronic trade actions the price of a commodity changes dramatically, then this may have consequences for people, say farmers and consumers, in places that are very remote from trading centers such as London, New York, or Tokyo. This epistemic invisibility, which becomes a moral invisibility, renders it difficult for traders to exercise responsibility – and difficult for others to hold them responsible. But we, as citizens, also have difficulties to exercise responsibility, since we do not know much about these highly technological financial activities that happen in remote and lofty financial centers and are alienated – or so it seems – from our daily lives. We do not know "who" and we do not know "where" since the humans and places are hidden from our sight. We only see "the market".

4. "THE MARKET" AS GHOST IN THE MACHINE

Indeed, the entire finance system can be seen as a large technological machine in which "markets" function as artificial agents. It is often said that "the market" "does" or "thinks" this and that. It is seen as an agent in its own right, as a 'greater being' as one of Knorr Cetina calls it [6]. Markets thus function as a kind of ghosts in the machine or gods from the machine (deus ex machine). They are transcendent gods since they become distant from the concrete, earthly human world. They are ghosts since they are forms reminiscent of humans, but are no longer human: in the global world of technofinance, electronic technologies contribute to the disappearance of the humans "behind" the markets as they abstract from concrete humans and social relations. Money (exchange) does this already to some extent, as Simmel has argued [9]; but electronic forms of money and trade increase the distance.

Again this ethically problematic, since "the market" cannot be a responsible agent. Ghosts or transcendent gods do not fulfil the criteria of moral agency and responsibility; only humans do. Transcendent gods are too distant too mingle with the world and ghosts are not supposed to act (in the world). Similarly, "the system" or "the computer" cannot be held responsible. The result is that the people who make trading decisions and exchange goods remain out of sight and can therefore escape responsibility ascription. Those who produce "the market" escape democratic control.

5. REVEALING ALTERNATIVE FINANCIAL-TECHNOLOGICAL PRACTICES

How can we change this? How can we decrease the distance? How can we increase visibility of humans and places? How can we render finance more socially and ethically responsible and what does this mean for responsible innovation? At first sight, it seems that there is not much we can do. But this view is mistaken. As the mentioned literature in social studies of finance and STS also shows, finance remains a human practice, and humans can change technologies. If we understand finance as a social and technological-human practice, this opens the way to changing society through finance. The world of finance may seem distant, but the ghosts, gods, and artificial agents are created by humans and their actions and thinking. Resistance is possible, and alternative practices are also possible. Yet often these alternative practices remain hidden, since they often emerge bottom-up, at grassroots level (communities, groups of individuals, small companies), rather than top-down (the world of governments and large financial institutions and corporations). Consider for instance so-called LETS (Local Exchange Trading Systems) and new "virtual" currencies: they have not been introduced by governments or national banks, but have been invented and experimented with by local communities or have emerged from internet-based, non-governmental initiatives and groups. But such financial technologies and financial practices are generally less well-known and are not so prominently present in the (mainstream) media as one may expect given their novelty and real-world impact. Thus, here we encounter another kind of morally problematic invisibility: the hidden efforts of many individuals and communities to re-invent finance, including the often unseen development of new, alternative social-financial practices and technologies. If people have the idea that change is not possible, it is difficult to bring about social-political change.

In response to this invisibility, then, we (as researchers) can help to reveal that there is already social-financial-technological change. Consider alternative trade systems such as fair trade, organic food systems and farmers markets, barter networks, LETS (again), time banks, and perhaps also "virtual" currencies in games or new electronic currencies such as Bitcoin which enable peer-to-peer exchange. There are microcredits and there is web-based peer-to-peer lending, for instance the platform Kiva. Gaming is also a way to explore different kinds of ways to deal with money and exchange. There are movements such as Slow Money and Positive Money. There is room for change: top-down but also bottom-up; there are many grassroots initiatives which do not only show new financial technologies but also *that* alternatives are possible, that change is possible – even and also in finance.

6. CONCLUSIONS FOR ETHICS OF FINANCE AND RESPONSIBLE INNOVATION

Ethics of finance – like most thinking categorized under this term – is usually concerned with principles that are to guide human action, or with human institutions and human character (e.g. virtue). But given rapid technological change in finance and elsewhere in our societies, and given research that reveals that these technologies have a significant impact on our lives and our world, such a limitation is highly problematic and undesirable. In order to repair this blink spot, ethics of finance needs to connect to thinking about technology – especially ICTs – and their ethical and social consequences. Thinking about technology (including computer ethics, information ethics, etc.) can in turn can learn from social studies of finance that reveal the technologies of finance and how they shape human experience, action, and society. Attention to invisible financial technologies and fictions such as "the market" may help us to think about how to render financial practices more responsible and how we, as citizens, can exercise our democratic responsibilities towards global finance. It is also important to show that there are already a lot of efforts to develop alternative financial practices and technologies, and to explore how these efforts can be supported. we need to discuss how we can create a society in which this kind of innovation becomes *more* visible and can draw on more resources than it presently does. We need to think about how to create environments in which experiments with financial-social change can flourish.

This requires re-thinking innovation but also responsible innovation. First, we must take a pro-active approach and connect ethics with actual process of innovation. If it is true that financial technologies are not mere instruments but also have social and ethical consequences, as this paper has suggested, then making finance more responsible should include making technological innovation in the context of finance more responsible. Can we design better financial technologies that support rather than threaten the exercise and ascription of responsibility? Can we use information and communication technologies in ways that bring about social change? As said, there are already interesting initiatives in this area. And of course there is already work on responsible innovation in other areas of technological development, which may be applied to financial innovation. Second, however, for this purpose the scope of thinking about responsible innovation must be broadened: not only in terms of the kind of technology (financial technologies), but also in terms of the sites and places of innovation. Usually thinking about responsible innovation has a rather narrow scope: it tends to focus on "the usual suspects": innovation in academia and industry. This is not sufficient, since it hides what happens outside of these domains. I therefore recommend that the scope of thinking about responsible innovation be widened to include grassroots, small-scale, and non-governmental innovation.

Finally, the approach used in this paper, which develops and employs what we may call a phenomenology of (in)visibility and which is informed by work on technologies by social scientists (especially researchers using ethnographical methods to study financial practices), could be applied more widely in ethics of technology and thinking about so-called research and responsible innovation (RRI). (See for instance Von Schomberg's definition [10].) Part of what ethics of technology and RRI is then supposed to do is making visible technologies and the hidden ways they shape our lives and society. By firmly connecting the normative to the descriptive, and ethical questions to epistemological questions, the proposed approach may thus contribute to a more engaged and socially responsive way of doing ethics of technology and RRI. Indeed, this paper suggests that to make visible these "robots", machines, people, and places – with the help of social studies or indeed by doing these studies – is already an ethical, responsible act; it helps to bridge the distance.

7. REFERENCES

[1] Boatright, J. R. (ed.) 2010. Finance Ethics: Critical Issues in Financial Theory and Practice. Hoboken, New Jersey: John Wiley & Sons.

[2] Heidegger, M. 1927. *Being and Time* (trans. J. Stambaugh). Albany, NY: SUNY Press, 1996.

[3] Harford, T. 2012. High-frequency trading and the $440m mistake. BBC News Magazine. Retrieved from http://www.bbc.co.uk/news/magazine-19214294

[4] Coeckelbergh, M. 2015. *Money Machines: Electronic Financial Technologies, Distancing, and Responsibility in Global Finance*. Ashgate, Farnham, UK.

[5] Callon, M., Ed. 1998. *The Laws of the Markets*. Blackwell, Oxford.

[6] Knorr Cetina, K. 2005. From pipes to scopes: The flow architecture of financial markets. In *The Technological Economy,* A. Barry and D. Slater (eds). Routledge, London.

[7] Beunza, D. and Stark, D. 2004. Tools of the trade: The socio-technology of arbitrage in a Wall Street trading room. *Industrial and Corporate Change* 13, 2, 369-400.

[8] Aristotle. *Nicomachean Ethics*. In The Complete Works of Aristotle. Vol II (ed. J. Barnes). Princeton: Princeton University Press, 1984.

[9] Simmel, G. 1907. *The Philosophy of Money*, ed. D. Frisby trans. T. Bottomore & D. Frisby. Routledge, London and New York, 2004.

[10] Von Schomberg, R. 2011. Prospects for Technology Assessment in a framework of responsible research and innovation. In: *Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methode*, pp.39-61, Wiesbaden: Springer VS.

The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots

Kathleen Richardson
CCSR
De Montfort University
Leicester, UK
44 (0) 116 2078584
Kathleen.richardson@dmu.ac.uk

ABSTRACT
In this paper I examine the model of asymmetrical 'relationship' that is imported from prostitution-client sex work to human-robot sex. Specifically, I address the arguments proposed by David Levy who identifies prostitution/sex work as a model that can be imported into human-robot sex relations. I draw on literature in anthropology that deals with the anthropomorphism of nonhuman things and the way that *things* reflect back to us gendered notions of sexuality. In the final part of the paper I propose that prostitution is no ordinary activity and relies on the ability to use a person as a thing and this is why parallels between sex robots and prostitution are so frequently found by their advocates.

Categories and Subject Descriptors
K.4.1 [Computers and Society]: Ethics

General Terms
Human Factors.

Keywords
Robot Ethics, Sex Robots, Prostitution, Subjectivity, Gender

1. INTRODUCTION

A number of initiatives are now in place to incorporate the development of sex robots into mainstream robotic activity. For example, in November 2015, roboticists interested in developing the area of sex robots can participate in the Second International Conference on Love and Sex with Robots to be held in Malaysia. The conference will explore topics such robot emotions, humanoid robots, teledildonics, and intelligent electronic sex hardware.

In his book, *Sex, Love and Robots* [1] David Levy proposes a future of human-robot relations based on the kinds of exchanges that take place in the prostitution industry. Levy explicitly creates' 'parallels between paying human prostitutes and purchasing sex robots' [1 p.194] . I want to argue that Levy's proposal shows a number of problems, firstly his understanding of what prostitution is and secondly, by drawing on prostitution as the model for human-robot sexual relations, Levy shows that the sellers of sex are seen by the buyers of sex as *things* and not recognised as human subjects. This legitimates a dangerous mode of existence where humans can move about in relations with other humans but not recognise them as human subjects in their own right.

What are the ethics of extending robots into new fields such as sex and what model of sexual relationship is invoked in the transference to robots? Ethically, there is a strong reaction to the use of robots in the military, and as such a well established organisation The Campaign to Stop Killer Robots (http://www.stopkillerrobots.org/) is devoted to preventing automated and robotic warfare developments that further take humans out of the loop. Should we as a robotic community also reflect on implementing a similar response to the development of sex robots? Could the development of sex robots also mark a disturbing trend in robotics? I will propose at the end of this article the urgent need to establish a Campaign Against Sex Robots.

2. CONSUMPTION OF INTIMATE BODIES AS 'GOODS'

Prostitution is the practice of selling a sex for monetary payment. In recent years those who work in the prostitution industry (particularly in Europe and North America) have promoted the term 'sex-work' over prostitution as a way to show how it is similar to other kinds of service labour. A term like prostitution implies that the provider is in a subservient position. *Third Wave Feminism* proposes that women are not subservient but are making conscious choices to choose work that is influenced by their sex [2]. By contrast, the term 'sex-worker' extends the framework of labour to include sexual work. This redefinition of prostitution to sex-work (and therefore framed as a service) has been challenged by a number of campaigners and scholars [3, 4, 5]. While those in favour of the sex industry describe it as an extension of free sexual relations, campaigners against prostitution point to the fact that in the absence of consent, prostitution cannot be reframed as positive. The facts of prostitution are disturbing where violence and human trafficking are frequently interconnected [3, 4]. Moreover the industry is extensive and a recent European Union Survey found:

-prostitution revenue can be estimated at around $186.00 billion per year worldwide.

-prostitution has a global dimension, involving around 40-42 million people worldwide, of

-which 90% are dependent on a procurer. 75% of them are between 13 and 25 years old.

[4 p. 6]

When robots are introduced as possible alternatives to women (or children), some, like Levy ask 'what's the harm? It's only a machine?' The same views are also proposed by some towards those who sell sex.

Levy also proposes that sex robots could help to reduce prostitution. However, studies have found that the introduction of new technology supports and contributes to the expansion of the sex industry. There are more women are employed by the sex industry than any other time in history [5]. Prostitution and pornography production also rises with the growth of the internet. In 1990, 5.6 per cent of men reported paying for sex in their lifetime, by 2000, this had increased to 8.8 per cent. These figures are likely to be even higher due to the reluctance of people admitting to paying for sex [6]. As the buying of sex relies on only acknowledging the needs of the buyer, it is no surprise that children also suffer as a consequence. The National Crime Agency in the UK has identified the web as a new source of threat to children including the proliferation of indecent images of children and online child sexual exploitation [7].

The arguments that sex robots will provide artificial sexual substitutes and reduce the purchase of sex by buyers is not borne out by evidence. There are numerous sexual artificial substitutes already available, RealDolls, vibrators, blow-up dolls etc., If an artificial substitute reduced the need to buy sex, there would be a reduction in prostitution but no such correlation is found. To understand why males buy sex it is important to understand what happens in an exchange and how males describe what is happening. The following are statements from males who buy sex:

'Prostitution is like masturbating without having to use your hand',

'It's like renting a girlfriend or wife. You get to choose like a catalogue',

'I feel sorry for these girls but this is what I want' [3 p.8].

While males are the chief buyer of human sex, females are more likely to purchase artificial nonhuman substitutes such as vibrators [1] that stimulate a discrete part of the body rather than purchase an adult or child for sex. Take a look again at the above– 'renting a girlfriend' or 'feeling sorry for the girls' these and many more indicate that the buyer of sex is putting his needs over and above the other person. In the prostitution/client exchange both enter the encounter in specific ways. A study by Coy [3 p. 18] found the asymmetrical form of encounter between buyers and sellers of sex. As modern subjects, male and females have equal rights under the law, and these rights recognise them as human agents. In prostitution, only the buyer of sex is attributed subjectivity, the seller of sex is reduced to a *thing*. This is played out in multiple ways where…

> …a denial of subjectivity occurs when the experiences and feelings of the "object" are not recognised. This denial of women's subjectivity can also be understood as sexual objectification. Both were evident in these men's lack of empathy with the feelings of women in prostitution. They constructed her in their own minds, according to their own masturbatory fantasies, as opposed to recognising the reality of the woman's feelings. It is also telling that often the men switched from understanding the woman's situation and feelings to attributing to her what they wanted her to feel during or after sex [3 p. 18].

In the sex exchange in prostitution, the subjectivity of the seller of sex is diminished and the subjectivity of the buyer is the only privileged perspective and viewpoint. As robots are programmable entities with no autonomous (or very limited) capabilities, it seems logical then that prostitution becomes the model for Levy's human-robot sex relations.

A key factor that is missing is the inability of the buyer of sex to have empathy with the seller of sex. Expert of autism, Simon Baron-Cohen [8] in his book *Zero Degrees of Empathy* proposes a gendered basis to empathy as a normative category. Baron-Cohen has this to say about empathy:

> Empathy is without question an important ability. It allows us to tune into how someone else is feeling, or what they might be thinking. Empathy allows us to understand the intentions of others, predict their behavior, and experience an emotion triggered by their emotion. In short, empathy allows us to interact effectively in the social world. It is also the "glue" of the social world, drawing us to help others and stopping us from hurting others [9 p.163].

Baron-Cohen suggests that the higher prevalence in crime, sexual abuse, the use of prostitutes and murder are disproportionately committed by men and show that men lack empathy in comparison to females [8]. By proposing that empathy is an ability to recognise, take into account and respond to another person's genuine thoughts and feelings is something that is absent in the buying of sex. The buyer of sex is at liberty to ignore the state of the other person as a human subject who is turned into a thing.

3. 'DOWNLOADING' HUMAN LIFWORLDS INTO THINGS

The use of robots for sex (adults and children) are justified on the basis that robots are not real entities, they are things. This narrative is also replayed in the production of video nasties, sexual abuse images of children in virtual reality settings [11] and the sexual and racial violence seen in some video games such as Grand Theft Auto where gamers are rewarded for killing prostitutes [12]. The transference of humanlike qualities to things has provoked extensive discussion in the robotics community. Is it possible to transfer human constructs of gender, class, race or sexuality to a robot or nonhuman? Anthropologically speaking the answer is yes. This theme has been replaced in a discussion of robots as slaves. Bryson [10] has railed against arguments associating robots with slaves because, she argues, they are nothing more than mechanical appliances –do to robots what you wish. But is it only possible to have an either or position? Is it possible then to propose that sex robots are harmful, knowing they are not human? While Bryon has important arguments, the way that human attribute meanings to robots, nature and animals reflect back to us what is of value.

But where do the fantasy images and products come from? Is fantasy just a neutral domain that is a sphere separated off from the 'real' and therefore unproblematic? I propose that fantasy, and the ways that robots are seen show human relations at work. The question is not do humans extend their lifeworlds into robots but what is being transferred to the robot? Anthropologists have developed an extensive literature on the anthropomorphism of things, framing it within the context of 'animism' as the attribution of a spirit to nonhuman animals [13, 14]. Moreover, the anthropology of technology explores how gender, class, sexuality and race is inflected in the cultural production of technological artefacts [15, 16, 17]. In a forthcoming paper I propose that technological-animism is at work in the sphere of robotics, but rather than come from spirit or religion as in classical studies, technological-animism comes from a lack of awareness and attention given to how cultural models of race, class and gender are inflected in the design of robots [18]. The issue then becomes not a why question (that is still open for debate), but a how question. In what ways are robots made and what uses are they put to and what can these practices tell us about gender, power, inequalities, race and class? Campaigns to extend rights to robots without due attention paid to human are problematic. Robertson [19] notes that campaigns to extend rights to robots are done in contexts where the campaigners do not simultaneously campaign for the extension of rights to all human beings. When this happens it is important to explore the ethics of the human that is reproduced in robotics. In some cases, such as sex-robots it will rest on a disturbing vision of a seller of sex as a thing.

In a recent article on gender and robots, Watercutter [20] highlighted the recurring imagery in fictions and robotic labs which overly presented female robots as young, attractive and focused on performing roles in the service industry as receptionists or waitresses. When it come to the explicit design of sex robots, Roxxxy designed by New Jersey-based company TrueCompanion shows a male view of a sexually attractive adult female complete with three points of entry in the body, the mouth, the anus and the vagina. But the development of sex robots is not confined to adult females, adult males are also a potential market for homosexual males. But the potential for a market in sex robots will be extended to child sex robots. Some researchers such as Ronald Arkin, professor of mobile robotics at Georgia Institute of Technology proposed that child robots could also be used in the treatment of paedophilia [21].

4. CAMPAIGNS AND ROBOTS

In this paper I have tried to show the explicit connections between prostitution and the development and imagination of human-sex robot relations. I propose that extending relations of prostitution into machines is neither ethical, nor is it safe. If anything the development of sex robots will further reinforce relations of power that do not recognise both parties as human subjects. Only the buyer of sex is recognised as a subject, the seller of sex (and by virtue the sex-robot) is merely *a thing* to have sex with. As Baron-Cohen shows, empathy is an important human quality. The structure of prostitution encourages empathy to be effectively 'turned-off'. Following in the footsteps of ethical robot campaigns, I propose to launch a campaign against sex robots, so that issues in prostitution can be discussed more widely in the field of robotics. I have to tried to show how human lifeworlds of gender and sexuality are inflected in making of sex robots, and that these robots will contribute to gendered inequalities found in the sex industry. I did not create these parallels between prostitution and the making of sex robots, these have been cultivated and explicitly promoted by Levy [1]. By campaigning against sex robots, we will also promote a discussion about the ethics of gender and sex in robotics and help to draw attention to the serious issues faced by those in prostitution.

5. ACKNOWLEDGMENTS

My thanks to the Centre for Computing and Social Responsibility (CCSR).

6. REFERENCES

[1] Levy, D. (2009). Love and sex with robots: The evolution of human-robot relationships. New York.

[2] Synder-Hall, C, 2010 'Third-Wave Feminism and the Defense of "Choice"' Perspectives on Politics, Vol. 8, No. 1 (March 2010), pp. 255-261.

[3] Farley, M., Bindel, J., & Golding, J. M. 2009. Men who buy sex: Who they buy and what they know (pp. 15-17). London: Eaves.

[4] Schulze, E, Novo, S.I, Mason, P, and Skalin, M. 2014 Research Assistant Gender Exploitation and Prostitution and its impact on Gender Equality. Directorate General for Internal Policies. http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/493040/IPOL-FEMM_ET(2014)493040_EN.pdf

[5] Barton, B. 2006. Stripped: Inside the lives of exotic dancers. NYU Press.

[6] Balfour, R . and Allen,. 2014. A Review of the Literature on Sex Workers and Social Exclusion. By the UCL Institute of Health Equity for Inclusion Health, Department of Health. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/303927/A_Review_of_the_Literature_on_sex_workers_and_social_exclusion.pdf

[7] National Crime Agency, *Child Exploitation* http://www.nationalcrimeagency.gov.uk/crime-threats/child-exploitation

[8] Baron-Cohen, S. 2011. Zero degrees of empathy: A new theory of human cruelty. Penguin UK.

[9] Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. Journal of autism and developmental disorders, 34(2), 163-175.

[10] Bryon, J.B., Close Engagements with Artificial Companions: Key social, psychological, ethical and design issue, Yorick Wilks (ed.), John Benjamins (chapter 11, pp 63-74) 2010.

[11] Eneman, M., Gillespie, A. A., & Stahl, B. C. 2009. Criminalising fantasies: The regulation of virtual child pornography. In Proceedings of the 17th European Conference on Information Systems (pp. 8-10). Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[12] DeVane, B., & Squire, K. D. 2008. The meaning of race and violence in Grand Theft Auto San Andreas. Games and Culture, 3(3-4), 264-285.

[13] Swancutt, K and Mazard, M (eds.) forthcoming 2016. Special Issue: The Anthropology of Anthropology in Animistic Ontologies. *Social Analysis.*

[14] Boyer, P. 1996. What makes anthropomorphism natural: Intuitive ontology and cultural representations. Journal of the Royal Anthropological Institute, 83-97.

[15] Helmreich, S. 1998. Silicon second nature: Culturing artificial life in a digital world. Univ of California Press.

[16] Martin, E. 2001. The woman in the body: A cultural analysis of reproduction. Beacon Press.

[17] Richardson, K. (2015). An Anthropology of Robots and AI: Annihilation Anxiety and Machines. Routledge, New York.

[18] Richardson, K. Forthcoming 2016. Technological Animism: The Uncanny Personhood of Humanoid Machines. In Swancutt, K and Mazard, M (eds.) forthcoming 2016. Special Issue: The Anthropology of Anthropology in Animistic Ontologies. Social Analysis.

[19] Robertson, J. (2014). Human rights vs. robot rights: Forecasts from Japan. Critical Asian Studies, 46(4), 571-598.

[20] Watercutter, A. 2015. Ex Machina has a serious Fembot problem. Wired. 9th April 2015. http://www.wired.com/2015/04/ex-machina-turing-bechdel-test/

[21] Robertson, J. (2010). Gendering humanoid robots: robo-sexism in Japan. Body & Society, 16(2), 1-36.

[22] Hill, K, 2014. Are Child Sex Robots Inevitable? Forbes Magazine.http://www.forbes.com/sites/kashmirhill/2014/07/14/are-child-sex-robots-inevitable/

Addressing Responsible Research and Innovation to Industry – Introduction of a Conceptual Framework

Emad Yaghmaei
Mads Clausen Institute
Alsion 2, DK-6400 Sønderborg, Denmark
+45 6550 9382
emad@mci.sdu.dk

ABSTRACT
Responsible research and innovation (RRI) is taking a role to assist all types of stakeholders including industry to move research and innovation initiatives to responsible manner for tackling grand challenges. The literature on RRI focuses little on how industry can implement RRI principles. In solving such gap in the literature, this article constructs a solid framework that provides a conceptual starting point for future research on levels of RRI. It draws a fundamental path to align industrial activities with environmental and societal needs. The framework develops a normatively grounded conceptual path for managing and assessing RRI principles in industry. This study depicts five successive RRI implementation stages and exhibits three RRI dimensions that represent different categories and corresponding indicators for that. The rationale behind this framework has been derived from extant models of corporate social responsibility (CSR) literature. Drawing on these models, this study develops stages and dimensions of RRI for discussing why industry should become engaged in RRI, how industry can embed RRI principles into research and innovation processes, how companies progress from one RRI stage to another, and how industry can manage all RRI dimensions systematically.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Ethics

General Terms
Management

Keywords
RRI in industry, RRI awareness, RRI assessment, RRI implementation, RRI stages, RRI dimensions

1. INTRODUCTION
Responsible research and innovation (RRI) has emerged in recent years as a potential bridge between science and society that aims to increase the public value of science. The European research framework Horizon 2020 has also a dedicated section for RRI entitled 'Science with and for Society'. In a nutshell, RRI is about better aligning the needs and values of society with what is happening in the world of science. When this alignment is not well done, the outcomes of research and innovation (R&I) tend to lose their legitimacy. RRI is to prove legitimacy of R&I, and from a holistic view, prove legitimacy of science. In essence, modern societies increasingly rely on research and innovation (R&I) to address the most pressing worldwide problems such as demographic change, security, and environmental or social sustainability. Current European policy specifically underlines the importance of R&I in addressing these so-called "grand challenges" and, more generally, to tackle them promoting a responsible approach to R&I. [7, 9, 10]

Addressing the grand challenges successfully would lead to the prospect of living a safe life with increased quality of life [14]. As a result of grand challenges, various social, cultural, economic and environmental problems, such as sustainable energy, affordable health care, cyber security, economic wellbeing and growth, demographic change, and child mortality, have emerged in the globe. RRI implementations plan for industry could address such regulatory gaps and tackle existing grand challenges.

The present study aims to contribute to RRI literature by establishing a conceptual framework, shows different stages of RRI in industry in connecting to RRI dimensions; In essence, this paper investigates a progressive integration of societal concerns into firms' management process, renders a theoretically robust basis for delineating responsibility trajectory–from to higher stages of RRI-, and propose a conceptual framework that operationalize RRI in industry.

To address these aims, the conceptual framework is developed as the basis for discussing why RRI components should be integrated into industrial levels, how they might be assess internally and systematically, and how they can implement in industry.. The literature on RRI is reviewed in the next section. This article will be continued by an overview of for managing RRI in industry by shedding light on RRI stages and RRI dimensions.. At the end, this study discusses about potential connections between RRI stages and RRI dimensions and proposes some for future research.

2. RESPONSIBLE RESEARCH AND INNOVATION
In defining RRI, Von Schomberg [30, p. 9] argues that "Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding

Conference'10, Month 1–2, 2010, City, State, Country.

of scientific and technological advances in our society)." Von Schomberg's definition of RRI, the most commonly cited in the literature, is broad and underscores several important aspects. The concept has common currency for at least three reasons. First, it reflects RRI concerns, where the process and product of innovation must be taken into account and RRI activities aim at ethical acceptability and societal desirability of R&I outcomes. Second, public engagement of different stakeholders (RRI actors) is regarded crucial ingredient for RRI. Third, the need of having improved outcomes of research and innovation is well captured in the description of RRI (RRI norms/values). The extant literature presents the concept of RRI in quite similar ways; from Stahl point of view [26], RRI is regarded as a higher-level responsibility or meta-responsibility by which R&I outcomes must be enriched through shaping, maintaining, developing, coordinating and aligning existing activities, actors and normative expectations [4]. This idea is represented in the following figure:

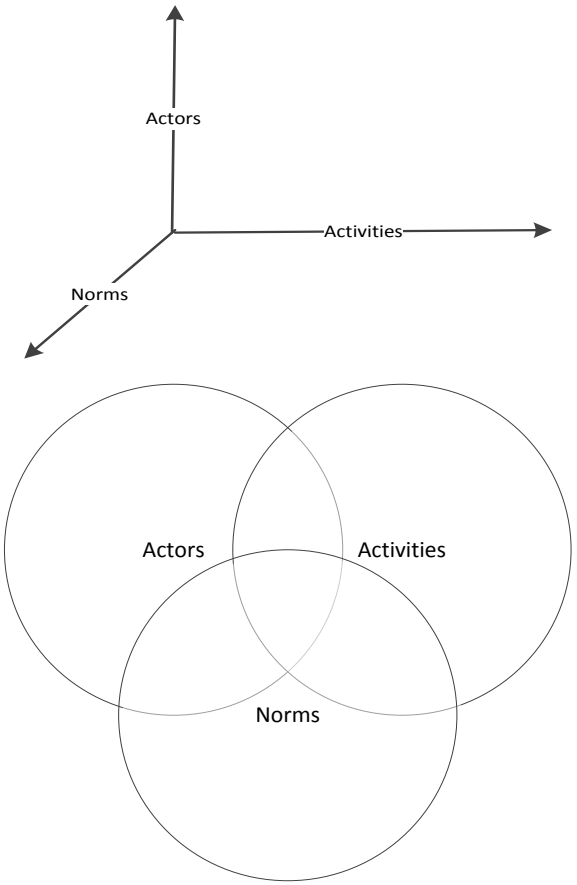


Figure 1. Different Attempts to represent the Space of RRI Graphically

Source: Stahl (2013)

2.1 RRI Actors

According to Figure 1, given engagement of various actors at different levels in the research and innovation systems, all existing RRI stakeholders should be addressed to align technology outcomes to values, needs and expectations of society [11]. These actors include policy-makers at national, regional and local level, professional bodies, legislators, research funders, individual researchers, research organizations (both publicly and privately funded), educational organizations, industry, users of research and innovation, research ethics committees and their members, civil society actors, and public bodies at different levels. This list is not exhaustive, but reflects general aim to shed light on the relevant actors.

In fact, there are substantial variations in the degree to which actors adopt RRI to become co-responsible for the innovation process. And the degree of responsibility/co-responsibility, which varies on the society, induced from a complex relationship between RRI dimensions: actors, activities, and values/norms.

2.2 RRI Activities

The constant development of RRI concept is derived when one may address a broad array of RRI activities, which the extant literature points them inclusively. In fact, RRI actors may use activities of RRI, develop existing RRI governance practices, perceive plausible regulatory gaps in relevant activities, and announce their needs for further initiatives to apply for better aligning the needs and values of society with what is happening in the world of science. Despite the fact that the present paper does not have enough space to cover these RRI activities fully, but in order to ahead sections become more touchy and effective for readers, an overview of these activities is briefly described.

EC's (European Commission) "Science and Society" (SaS), "Science in society" (SiS), and recently developed "Science with and for society" themes all acquired a strong focus on six action lines, which addressed as central policy priorities for RRI. In essence, EC has decided to include RRI as a cross-cutting aspect in the implementation of the new European Research Framework Programme Horizon 2020. These EU focus areas include societal engagement (better engagement of citizens to science), gender equality (enhanced presence of women in science), science education (improved science literacy and education of all Europeans), open access (e.g. open access to scientific results), ethics and governance (better aligned, responsible and more efficient governance of science). [28]

To become more responsive to society's needs through R&I system and aligning technology outcomes with and for the needs and values of society, R&I projects need to be assessed if they are socially and ethically desirable and acceptable. Many ways of assessing aspects of R&I projects have been identified over time; there are including risk assessment, impacts assessment, and technology assessment. Kermisch [16] argues an integral connectivity between risk and responsibility in which the need of integrating RRI into research and innovation could be fulfilled by risk assessment [21]. Another type of assessment is impact assessment in which identifications of possible consequences of particular types of risk, in turn subsequent assessing would perform proactively. Privacy impact assessments (PIA) [2, 15] and ethics impact assessment based on Wright model [32] reflect unequivocal role of impact assessment in RRI. Furthermore, in covering technology assessment, Grunwald points that "technology assessment (TA) as a most common collective designation of the systematic methods used to scientifically investigate the conditions for and the consequences of technology to denote their societal evaluation" [13]. TA shall add reflexivity to technology governance [1] by integrating any available knowledge on possible side effects, by supporting the evaluation of technologies according to societal values and ethical principles, by elaborating strategies to deal with inevitable uncertainties, and

by contributing to constructive solutions of societal conflicts around science and technology.

In view of the complexity of relations between the needs and values of society with what is happening in the world of science, prospective studies and foresight activities contribute to RRI-related activities assessment to focus on the consequences of science, likewise address grand challenges. As indicated in the Owen work [22], RRI aims to be anticipatory by using foresight techniques. In essence, while lack of technology foresight is labeled as one single irresponsible actor [31], future studies [24] or foresight research [3, 20] can be identified as one single responsible actor. However, as many actors are involved in innovation processes, neither irresponsible nor responsible outcomes are seldom the result of one single actor.

Allocating roles of responsibility to all stakeholders engaged in the research and innovation processes is a key component of RRI, which induced from leading responsible governance models [29]. In essence, due to the need for the legitimacy of research funding and certain scientific and technological advances, public engagement with science and technology across all involved actors should be taken into account [30]. In the presence of strong public engagement, at their different mechanism levels- upstream, midstream, and downstream [23], "technology push" and "policy pull" of new technologies are addressed and, ideally, would be moderated through involvement of actors.

To address other RRI action lines argued by EC including ethics, gender equality, open access, science education, and governance, they need to be drawn on a further range of processes and activities. One may argue integration of ethical values into research and innovation processes, which emerges, for instance, at value-sensitive design to facilitate such integration [19]. Likewise, responsible governance of research and innovation should be addressed from all stakeholders. Further, scientific education lies at the core of RRI activities by which increase the perceived value of RRI and awareness among stakeholders. As such, training approach on its various levels would find its linkage to RRI [12]. To operate RRI in proper way, stakeholders need to engage women in science, especially at senior levels to promote gender equality in science. In addition, returning to EC action lines and following an explicit policy goal, open access to scientific knowledge, research results and data is deemed as a basis term to boost innovation and increase the use of scientific results by all societal actors. Hence, the progress towards more open access in policy (member states level), strategy (research organization level) and in performance should be taken into account.

There is further range of activities in the space of RRI such as external evaluation and professionalism [26], project reflexivity [30], RRI-sensitive research methodology [19], and standardization and regulation [6, 8]. Collectively, this enumeration of activities does not claim to be exhaustive, but all of activities fall under the term of RRI.

2.3 RRI Values and Norms

To create a responsible research and innovation process, a holistic view of the value proposition is required that covers the social, cultural, economical and environmental benefits of R&I. Research and innovation should responsible to the needs and expectation of society and reflect its values on different levels [5] Some of the central objectives are an improved quality of life and a reduction of the number of people living in poverty, an increased employment rate and employment opportunities for all citizens, respect for fundamental rights, sustainable development, a competitive social market economy, and respect for cultural and social diversity [11]. Hence, the consideration of value for RRI actors needs to be extended explicitly.

To summarize, RRI action lines involves actors taking responsible roles - address RRI activities - by spelling-out "regulatory gaps". To develop such engagement, having an RRI-based organizational learning is proposed in this article. Particularly this study seeks to develop a management approach for industry as one of RRI actors in which explain why industry should become involved in RRI, how industrial stakeholders in different sizes may implement RRI components internally and systematically, and how their progress can be assessed by internal staff.

3. MANAGING RRI IN INDUSTRY

Industry does not take the same actions for implementing RRI as other RRI actors do. One can argue the aim of RRI in industry is to ensure positive impacts of technology for exploring and capturing high level of responsibility in research and innovation (R&I) initiatives; In essence, new research, products and services should design and develop by set of functional requirements to address the grand challenges, reduce the regulatory gaps, obtain appropriate knowledge on the consequences of the outcomes of R&I, and evaluate both outcomes and options in terms of societal needs and moral values effectively and successfully.

Researchers have only recently focus on how RRI principles might be implemented in industry. As a result, little is known about components of RRI implementation plan for industry. What current RRI initiatives fail to emphasize to a sufficient degree is managing RRI principles in industry given industry' characteristics. To understand how RRI principles could integrate into industrial level, it is necessary to take into account awareness of RRI-related issues and convince industry to engage, map of a framework to company for assessing RRI-related issues, and implement RRI eventually. While couples of generic implementing tools have been identified in corporate social responsibility (CSR), few tools if any have been developed within RRI context to assist industry in the practical design of responsibility. Thus, this study firstly uses existing CSR tools to design a conceptual framework for various stages of implementation of RRI in industry, subsequently develop an RRI tool in better understanding responsible value creation within industrial stakeholders.

This paper identifies a need for a novel framework to assist industry in better aligning RRI principles along the value chain. This model seeks to align better six RRI action lines namely ethics; gender equality; open access; public engagement; science education; and governance with organizational practices and processes of companies. The ultimate aim of the research is to design a framework in which contributes to industry to implement RRI. To address this aim, this paper outlines the most important aspects of RRI in the extant literature; subsequently establish the conceptual framework.

Observing on collected data from RRI actors, the dimensions of RRI implementation plan have been shown at three levels, which are namely RRI awareness, RRI assessment, and RRI implementing. The three dimensions demonstrate the different types and corresponding indicators for RRI required. These dimensions depict how RRI is ideally understood and integrated within R&I practices and processes. In fact, they have been derived from raised assumptions in how industry can integrate RRI principles and methodologies into research and innovation processes. Hence, to integrate RRI, awareness of issues related to

RRI for convincing to engage, assessing of RRI besides mapping of a framework to company, and internal implementation of RRI should be taken into account.

To develop this schema of RRI more in depth and applicable for industry, this work also focuses on Zadek' model of CSR-based organizational learning [33]. This model represents successive stages of CSR implementation plan to integrate societal issues into organizational practices. For doing so in RRI context, this conceptual framework is inspired by Zadek' model from CSR in which show progressive stages of capturing RRI values from no accepting RRI principles in place to holistic RRI model with full implementation of RRI principles. In principle, the framework helps to link industrial stakeholders to issues related to RRI and observe societal concerns' integration steps in companies constantly. Collectively, this novel model aims to monitor RRI dimensions in different stages and see companies' progress in reducing grand challenges over time. Furthermore, this RRI framework offers industry a practical guideline tool for managing RRI by looking at defined RRI dimensions and RRI stages in industry.

4. FRAMEWORK

Three certain RRI themes emerged from the literature, which further acknowledge the need to design a framework for tracking RRI levels. They are namely RRI awareness, RRI assessment, and RRI implementation. At the same time, five common RRI stages – defensive, compliance, managerial, strategic, and civil- depict the need for embedding responsibility within research and innovation practices and processed in industry. This section draws a framework based on both above implications.

4.1 The Stages of RRI

One of the useful analytical tools in classifying companies behavior is stage models in which one may evaluate progressive steps of a certain behavior in companies over time [17, 18]. For RRI, the stages from lower to higher engagement in RRI are depicted in Figure 2. To render such a figure, applicable indicators are needed to set out how industry advances from a certain stage to another. These indicators will identify from responsibility report, code of conduct in firms, ethical reports, etc. The conceptual framework in this article aims to set of criteria, which classify the different stages of RRI. In this spirit, the Zadek' model [33] from CSR assists to draw this model. Zadek acquired five CSR stages namely defensive, compliance, managerial, strategic, and civil. In fact, to identify different levels of RRI principles, which integrated into research and innovation practices and processes, similar stages for RRI inspired by CSR are needed to show progressive steps of company involvement in RRI from defensive level to civil one. It is however fundamentally important to imply this point that whereas some R&I activities could be linked to more advanced stage like civil stage, other similar activities at the same company might not be advanced enough. In other words, RRI action lines have been integrated into some activities of a company more advanced rather into other activities. Thus, an overview of different stages determine the path to engage RRI action lines along the value chain for research and innovation activities.

At defensive stage, companies are not engaged in RRI activities, either because they are not aware of RRI action lines or deny social responsibilities in general, or due to not being able to address them. The adoption of compliance stage is more challenging for companies to which adhere to existing regulations which laid for social responsibilities, sustainability, and in particular RRI. In practice, however, while some operating regulations might be applied in distinctive areas, these laws do not follow in companies in other regions. Hence, the weakness of regulatory governance in few areas could be named as a significant problem in this stage. From managerial stage to the higher stages RRI principles are reflected into research and innovation practices; in essence, at managerial stage some issues related to RRI activities but not all of them, such as gender equality or ethics, is/are taken into account. Following the similar model, at strategic stage a firm has set a holistic RRI agenda to address full range of RRI principles by having particular protocols in place. At this stage, sometimes companies spending resources on reflecting RRI outcomes in their SWOT analysis. At the highest stage called civil stage, the company with the coherent RRI agenda in place, tries to promote RRI issues-related to third parties. In principle, the company action is leveraged to address, develop, and promote RRI agenda to others.

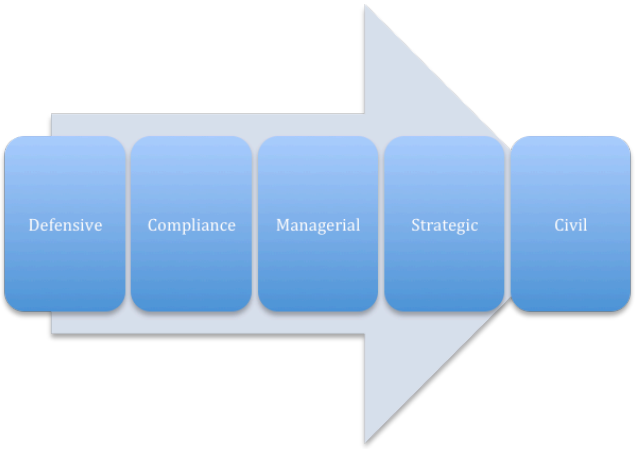


Figure 2. RRI implementation stages in industry

4.2 Dimensions of Implementing RRI

The three dimensions of this framework namely RRI awareness, RRI assessment, and RRI implementation are acquired from the RRI literature [4, 22, 26]. These RRI implementation dimensions help clearly in specifying RRI indicators in order for monitoring organizational characteristics in terms od responsibility; such indicators are named for instance acceptance of social interaction along value chain, acknowledgment of RRI principles in SWOT analysis, etc for RRI awareness. According to the logic of the study, industrial stakeholders firstly should become aware of issues related to RRI, subsequently acknowledge the social connection along the value chain, and eventually reflect RRI principles in their SWOT analysis. All these indicators in RRI-awareness dimension are included. After awareness level and based on established view of RRI, social concerns and democratic accountability should be integrated into R&I practices and processes so that the anticipatory governance principles for R&I is set. [27, 30]

The first dimension of "RRI awareness" indicates how companies might gain knowledge, access to, and work on integrating RRI into their processes. The purpose of this initial step is to convince top-level in companies for further discussion of RRI to find possible steps for implementation. Collectively with what explained about RRI stages so far, within the framework, companies awareness of action lines of RRI is varies from the lower stages to the higher ones; companies' prioritization for

addressing RRI action lines specify which certain RRI principles they are directly associated. For instance, a company needs to be evaluated in the case of ethics, or open access, and another company focus much more on gender equality and governance. As such, this dimension represent the extent to which a company is aware of issues related to RRI and accept a social interaction with other stakeholders along the value chain, and thus reflect RRI concerns into SWOT analysis. Hence, this dimension has three functional indicators: RRI-awareness, social-interaction acceptance along the value chain, and RRI principles acknowledgement in SWOT.

The second dimension namely RRI assessment is with a high level manager in the company who will be tasked with developing the strategy for implementing RRI. Managers carry different degree of RRI over the R&I processes. Thus, companies need to assess the role and distribution of RRI principles on R&I processes. In doing so, this framework refers to the literature where risk assessment, impact assessment, and technology assessment are the main directions of assessing of R&I processes [13, 21, 25, 26]. The risk assessment is fulfilled because of an integral connectivity between risk and responsibility [16]. Impact assessment in companies deal with consequences of particular types of risk, such as privacy impact assessments (PIA) and ethics impact assessment. Further, the third indicator as discussed above is technology assessment, which is a most common collective action. Technology assessment in this framework is in place to evaluate technologies in linkage with RRI action lines. As such, three operational indicators are determined for the RRI-related activities assessment: risk assessment, impact assessment, and technology assessment.

The third dimension is RRI implementation. One may assume that once the general agreement on RRI awareness and RRI assessment have been reached, the principles and ideas of RRI should be implemented; as such, the company should look at the actual RRI practices applying in the R&I processes. A concert RRI value proposition is developed when social, cultural, economical and environmental values are addressed correctly; therefore, industrial stakeholders must be connected to the other external actors in order for implementing RRI by fulfilling those values. Such an external connection initially reflects the extent to which company engage in RRI jointly with other actors and may be given on stakeholder involvement. Moreover, collective actions require to be existed in place to augment interactive activities between different RRI actors in implementing RRI activities. Thus, external RRI as a sub-dimension of RRI implementation dimension has its two operational indicators consist of collective action and stakeholder involvement.

Another sub-dimension of RRI implementation is identified in interaction with internal actors. Based on extant data in the literature, more internal detailed instances regarding RRI implementation can be studied such as the type of operational practices and procedures and staff involvement. In essence, the behavior of staff within RRI actors, in particular in industry could influence on R&I processes in which organizational governance structures are deployed for implementing RRI-related activities; furthermore, industry' operational practices and procedures could affect the scope of RRI-related tasks in industry. As such, at this dimension the interaction will be with an employee who undertakes R&I processes and the one takes care about integration of RRI into these processes. Hence, this sub-dimension is divided into two certain operational indicators including staff involvement and operational practices and procedures.

4.3 Stages and Dimensions

Based on these theoretical insights, the three dimensions and their ten operational indicators of the framework are interacted with five discussed RRI stages. The connection between these operational indicators and stages determines how industry could integrate principles and methodologies of RRI into R&I processes in industry. For now, it is not the purpose of this paper to go into much depth into elaboration of this connection; it is nevertheless important to know that this is a linkage between RRI stages and RRI dimensions.

This novel framework includes:

- Five stages of RRI awareness, which show the progressive steps from no awareness of RRI to full its awareness; companies may pass the stages from defensive reactions to civil level. Identifying stages separately assist in addressing RRI aspects within organizational practices and procedures.
- A segment to assess RRI activities within R&I processes. The proposed conceptual framework seeks to apply the range of assessments related to RRI such as risk assessment, impact assessment, and technology assessment. The level of relevant assessments coverage all stages including managerial and strategic ones.
- An implementation plan for RRI to capture RRI values and norms. The company is represented collective action and stakeholder involvement at external level as well as indicates operational practices and staff involvement at internal level in order for facilitating integration of RRI action lines along the value chain; this could happen in five consecutive stages, which inspired from CSR literature –defensive, compliance, managerial, strategic, and civil.

Collectively, by presenting a set of stages and dimensions for RRI, the conceptual framework is intended to map out all existing levels to deliver responsibility into industry.

5. CONCLUSION

Responsible research and innovation concept seems a key to manage future research and innovation processes to delivering meta-responsibility for society [4, 26]. The conceptual framework developed in this article is intended for embedding RRI action lines into R&I processes. In fact, this framework shows how companies move from one RRI stage to another. The aim of this article is to exhibit an ideal stage of RRI for industry. It is necessary in rising RRI awareness to meet responsible industry; although awareness is not sufficient criteria. In addition, it does require having some RRI assessments during research and innovation initiatives. Further, in implementing RRI action lines, the framework is considered to being applicable to industry.

Academia may add to the discussion of RRI stages and RRI dimensions by using this framework in industry empirically. This paper is just primary step in embedding RRI into the core of the company. Further work and more debate on social and ethical issues are recommended to optimize the framework by which assist industry to work productively together with RRI actors to achieve maximum impact of RRI.

6. REFERENCES

[1] Aichholzer, G., Bora, A., Bröchler, S., Decker, M., and Latzer, M. Ed. 2010. Technology Governance. *Der Beitrag der Technikfolgenabschätzung.* Berlin, Sigma.

[2] Clarke, R. 2009. Privacy impact assessment: Its origins and development. *Computer Law & Security Review,* 25, 2, 123–135. DOI= 10.1016/j.clsr.2009.02.002.

[3] Cuhls, K. 2003. From forecasting to foresight processes: New participative foresight activities in Germany. *Journal of Forecasting,* 22, 2-3, 93–111. DOI= 10.1002/for.848.

[4] Eden, G., Jirotka, M., and Stahl, B. 2013. Responsible research and innovation: Critical reflection into the potential social consequences of ICT. In *2013 IEEE Seventh International Conference on Research Challenges in Information Science (RCIS),* 1–12. DOI= 10.1109/RCIS.2013.6577706.

[5] Geoghegan-Quinn, M. 2012. Towards a European Model for Responsible Research and Innovation. *Conference "Science in Dialogue",* Odense, Denmark. (April 2012), 23-25.

[6] European Commission. 2009. Commission recommendation on: A code of conduct for responsible nanosciences and nanotechnologies research & Council conclusions on Responsible nanosciences and nanotechnologies research. DOI= http://ec.europa.eu/research/research-eu.

[7] European Commission. 2010. COM(2010) 2020: Europe 2020 - A strategy for smart, sustainable and inclusive growth. DOI= http://ec.europa.eu/eu2020.

[8] European Commission. 2011. A renewed EU strategy 2011-14 for Corporate Social Responsibility (No. COM(2011) 681 final). *Brussels: European Commission.* DOI= http://ec.europa.eu/enterprise/policies/sustainable-business/files/csr/new-csr/act_en.pdf.

[9] European Commission. 2012. Investing in Research and Innovation for Grand Challenges. *Brussels: European Commission, DG Research.* DOI= http://ec.europa.eu/research/erab/pdf/erab-study-grand-challanages-2012_en.pdf.

[10] European Commission. 2012. Responsible research and innovation - Europe's ability to respond to societal challenges. *The Directorate-General for Research and Innovation of the European Commission.* DOI= http://ec.europa.eu/research/science-society.

[11] European Commission. 2013. Chair: Jeroen van den Hoven, Options for Strengthening Responsible Research and Innovation. *Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation,* EUR25766 EN

[12] European Group on Ethics in Science and New Technologies. 2012. Ethics of Information and Communication Technologies (Opinion of the EGE No. 26, pp. 136). *Brussels: BEPA - Bureau of European Policy Advisors.* DOI= http://ec.europa.eu/bepa/european-group-ethics/docs/publications/ict_final_22_february-adopted.pdf

[13] Grunwald, A. 2009. Technology Assessment: Concept and Methods. In D. M.Gabbay, A. W. M. Meijers, J. Woods, and P. Thagard, Ed. *Philosophy of Technology and Engineering Sciences,* North Holland. 9, 1103–1146.

[14] Hinde, R. A. 2008. Bending the Rules: The Flexibility of Absolutes in Modern Life: The Twenty first Century Morality. *OUP Oxford.*

[15] Information Commissioner's Office. 2009. Privacy Impact Assessment Handbook, v. 2.0. DOI= http://www.ico.gov.uk/upload/documents/pia_handbook_html_v2/files/PIAhandbookV2.pdf.

[16] Kermisch, C. 2012. Risk and Responsibility: A Complex and Evolving Relationship. *Science and Engineering Ethics*, 18, 1, 91–102. DOI= 10.1007/s11948-010-9246-y.

[17] Kolk, A., and Mauser, A. 2002. The evolution of environmental management: From stage models to performance evaluation. *Business Strategy and the Environment,* 11, 14-31.

[18] Maon, F., Lindgreen, A., and Swaen, V. 2010. Organizational stages and cultural phases: A critical review and a consolidative model of corporate social responsibility development. *International Journal of Management Reviews,* 12, 20-38.

[19] Manders-Huits, N., and Van den Hoven, J. 2009. The need for a value-sensitive design of communication infrastructures. In P. Sollie, and M. Duwell, Ed. Evaluating New Technologies: Methodological Problems for the Ethical Assessment of Technology Developments, Heidelberg, Germany, *Springer,* 51–62.

[20] Martin, B. R. 2010. The origins of the concept of "foresight" in science and technology: An insider's perspective. *Technological Forecasting and Social Change,* 77, 1438–47. DOI= 10.1016/j.techfore.2010.06.009.

[21] Owen, R., and Goldberg, N. 2010. Responsible innovation: A pilot study with the U.K. Engineering and Physical Sciences Research Council. *Risk Analysis: An International Journal,* 30, 1699–707. DOI= 10.1111/j.1539-6924.2010.01517.x.

[22] Owen, R., Heintz, M., and Bessant, J. Ed. 2013. Responsible Innovation. *Wiley.*

[23] Rowe, G. and Frewer, L. J. 2005. A typology of public engagement mechanisms. *Science, Technology and Human Values,* 30, 251–90. DOI= 10.1177/0162243904271724.

[24] Sardar, Z. 2010. The namesake: Futures; futures studies; futurology; futuristic; foresight—What's in a name? *Futures,* 42, 177–84. DOI= 10.1016/j.futures.2009.11.001.

[25] Schot, J., and Rip, A. 1996. The past and future of constructive technology assessment. *Technological Forecasting and Social Change,* 54, 251–68. DOI= 10.1016/S0040-1625(96) 00180-1.

[26] Stahl, B. C. 2013. Responsible research and innovation: The role of privacy in an emerging framework. *Science and Public Policy*, sct067. DOI= 10.1093/scipol/sct067.

[27] Sutcliffe, H. 2011. A report on Responsible Research and Innovation. DOI= http://www.matterforall.org/pdf/RRI-Report2.pdf.

[28] Technopolis, & Fraunhofer ISI. 2012. Interim evaluation and assessment of future options for Science in Society Actions Assessment of future options. *Technopolis group.* Brighton, UK. DOI= http://ec.europa.eu/research/science-society/document_library/pdf_06/phase02-122012_en.pdf.

[29] Van den Hoven, J., Doorn, N., Swierstra, T., Koops, B. J., and Romijn, H. Ed. 2014. Responsible Innovation 1: Innovative Solutions for Global Issues. *Springer Dordrecht Heidelberg New York London.*

[30] Von Schomberg, R. Ed. 2011. Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields. *Luxembourg: Publication Office of the European Union.* DOI= http://ec.europa.eu/research/science-society/document_library/pdf_06/mep-rapport-2011_en.pdf.

[31] Von Schomberg, R. 2013. A vision of responsible innovation. In R. Owen, M. Heintz, and J. Bessant Ed. Responsible Innovation. London, *John Wiley, forthcoming.*

[32] Wright, D. 2011. A framework for the ethical impact assessment of information technology. *Ethics and Information Technology,* 13, 3, 199–226. DOI= 10.1007/s10676-010-9242-6.

[33] Zadek, S. 2004. The path to corporate social responsibility. *Harvard Business Review,* 82, 12, 125-132.

"Ask an Ethicist" – reflections on an engagement technique for industry

Catherine Flick
De Montfort University
The Gateway
Leicester, United Kingdom
+44 116 207 8487
cflick@dmu.ac.uk

ABSTRACT
In this paper, a method for engagement with industry stakeholders, namely, a booth at an industry convention entitled "Ask an Ethicist" is presented and reflected upon. Engagement methods included informal discussions with stakeholders, and challenges for attendees through targeted questions addressing ethical and social issues in the industry. While this booth was targeted at the video games industry, the author has reason to believe that it could be useful in other industry events to encourage and facilitate engagement between industry and society, and as a potential data collection tool for research into RRI in industry.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Ethics

General Terms
Human Factors

Keywords
responsible research and innovation; video games; ethics; stakeholder engagement; industry

1. INTRODUCTION
The video game industry is a multi-billion-dollar sector with a massive consumer base that is increasing every year. There has been significant criticism that the industry is, however, slow to move to deal with problematic social and ethical issues within it. In a previous paper [3] I have argued that to deal with these issues it is possible that Responsible Research and Innovation (RRI) could be a useful framework to allow industry to develop responsible video games that address, mitigate, or avoid these concerns. In that paper I also argued that academia is in a good position to be able to examine these issues sector-wide and contribute to the discussion of improvement of video games.

In the video game industry there are ideal places to engage with and collect opinions from a largely invested group of stakeholders – video game conventions. One such is PAX East, which has, over a three-day convention, over 50,000 attendees. It has, over the years, been a welcoming place for discussion and criticism of video games, through multiple talks given on social impact, and the inclusive community it has developed through its policies (such as anti-harassment and booth babe policies, which are not universal across video game conventions). In 2014 PAX East was the first video game convention to feature a "Roll for Diversity Hub and Lounge" which was developed to highlight equality and diversity issues within the video games industry and community. As a previous speaker on issues to do with ethics in video games, I applied for and was accepted for a booth entitled "Ask an Ethicist", with the aim to discuss ethical issues in video games with attendees.

This paper describes the two-way engagement nature and success of the engagement that was had through the booth and discusses the potential it could serve for further engagement with industry and stakeholders alike.

This paper also reflects on the "Ask an Ethicist" booth mechanism through which academics can become more involved in developing the discourse I advocated in my previous paper about responsible technologies, in this case video games. It looks at two years of engagement through this booth in the Diversity Lounge at PAX East. It also ponders whether the activities conducted at the booth could be useful for industry to determine the social desirability, acceptability, and other social and ethical issues to do with their upcoming video games, or whether such a booth could be a potential resource for video game developers to present questions of an ethical or responsible nature to stakeholders and provide meaningful data for industry to use.

2. ASK AN ETHICIST
This section explains the setting, setup, and methodology behind the "Ask an Ethicist" booth at PAX East.

2.1 Setting and Setup
PAX East [7] is a large video game convention aimed at video game players (unlike E3 which is aimed at industry and the press, mostly). It is held annually around Easter time in Boston, USA, and attracts well over 50,000 attendees during its 3-day weekend show. It is set up so that there is a main expo hall, where the video game companies (and related companies, e.g. computer hardware, streaming services, etc.) have areas to show off their games, halls for talks and panels, food areas, and a couple of specialist areas, including the "Roll for Diversity" Hub and Lounge (henceforth called the "Diversity Lounge"). The Diversity Lounge exists in a conference room and has booths for several diversity-interested groups, e.g. Toronto Gaymers, Northwest Press (a LGBT comic publisher), PressXY (a trans gamer group), AbleGamers (a group for disabled gamers), TakeThis (a mental health charity). I had heard that this room was going to be set up after it was reported in video game news, as a response to calls for more diversity in the video games industry (not without skepticism, however!) [5]. I had previously given a talk at PAX East 2013 on video games and

ethics, and had thought about trying to get a booth in the main expo hall, so I wrote to the organisers and asked if I could be involved. I suggested that a booth be set up with my expertise as the draw-card, an "Ask an Ethicist" booth, and was accepted for PAX East 2014 at the first ever Diversity Lounge. It was envisaged that this booth could be used to discuss ethical and social issues with attendees and to determine some important issues to them in order to foster further discourse and academic investigation of the issues.

2.1.1 PAX East 2014

In 2014 the booth was set up without much knowledge of booths and took quite a simplistic approach for a sign and some information with a cardboard poster and various prompt words around them (Figure 1).

Figure 1: PAX East 2014 "Ask an Ethicist" booth

On the table I had two pieces of flipchart paper with a "question for the day" on each. Participants could respond to the questions by writing answers on post-it notes and sticking them on the paper (Figure 2). The questions will be covered in more detail in the next section. I mostly staffed the booth, but I had a couple of friends helping me out to cover the booth while I had a break (we had a Snoopy-style "The ethicist is IN/OUT" sign as one of the posters). Finally, university signage and a bright green tablecloth with university logos on it made the booth more official-looking.

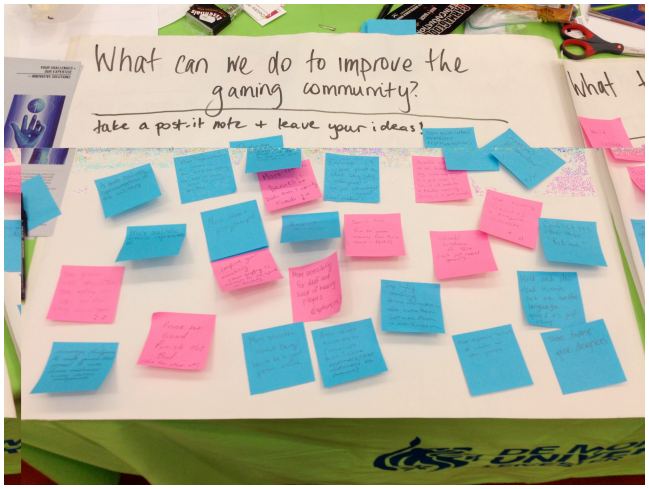


Figure 2: Post-it notes left to answer a question, PAX East 2014

2.1.2 PAX East 2015

In 2015 the booth was expanded to conduct some research through small interviews and targeted questions. More on the question choice and data collection will be discussed in the methodology section. The booth was much the same, except that instead of the "do-it-yourself" type signage, a professional sign was developed and printed. The "ethicist is in" aspect was dropped, as feedback from the previous year showed that there was less engagement when I wasn't there and it was obviously signed. As my helpers were well-instructed this year (and some had previous experience), this was considered unnecessary. The booth setup is shown in Figure 3.

Figure 3: PAX East 2015 "Ask an Ethicist" booth

This time I had an actual flip chart as well, so people could post the post-it notes up on the board. This allowed attendees to see a bit better what was happening with the post-its as in the previous year there really could only be 2-3 people close in to the booth reading and responding at a time. With the upright chart, it allowed larger groups of people to view (and contribute) at a time. The flipchart layout can be seen in Figure 4.

Figure 4: Flipchart layout, PAX East 2015

As can be seen, there were significant improvements from the first, pilot year, to the second. The upright poster made a big difference, as it looked a lot more professional. However, I did lose some of the visual interest from behind me – perhaps next year I could have some more "prompt" words attached to the drape.

It is important here to mention the general make-up of the people who visited the Diversity Lounge. While I have no official statistics, generally speaking the Lounge was likely to be a largely self-selecting space, with people interested in gender and diversity issues. There was not a lot of advertisement about it in the programme, for example, just a pop-up banner outside the room and a label on the convention centre map. In 2014, one of the founders of the convention, comic illustrator Mike Krahulik, who was partly responsible for the events that led up to the Diversity Lounge being set up, visited and that brought a large number of more general public into the lounge – people who wanted to see him and try his new game. In 2015 this didn't happen, but I had the impression there was around the same amount of traffic despite slightly fewer booths. If the Diversity Lounge had been on the show floor, it would have had far more foot traffic. However, one of the more valuable parts of the experience for me (and for participants) was the conversations I had with people about ethics and video games and their stories, which would have been virtually impossible to do in such a noisy atmosphere as the expo hall.

2.2 Methodology and Data Collection

This section discusses the methodology behind the data collection and analysis. As this paper will not be fully analyzing the data, but reflecting on the methods used, this section will be partly descriptive and partly reflective.

2.2.1 Questions at PAX East 2014

Initially PAX East 2014 was a pilot run, designed to try out a couple of question ideas to see what people would respond to. I had already ruled out doing a questionnaire or survey – I wanted the participants to feel as though they'd learned something from experiencing my booth as much as I had learned something from them. I'd seen the use of "short opinion post-it notes" in other scenarios, and thought it might work in this one. It was also an easy way to gather instant feedback on the question being asked.

I started out with some more "academic" questions – completely without realizing they were more academic: "What ethical issues are important in gaming?" and "What can we do to improve the games industry?". These proved to be problematic, and not particularly engaging, with only 17 responses for the former and 24 for the latter over a whole day. Many people had asked me "what is an ethicist?" so I realized that I would probably have to take the word "ethics" out of future questions and instead ask more relatable questions.

On the Saturday I asked "What do you do in video games that you don't do in real life?" and "What can we do to improve the gaming community?". The former question proved extremely popular with 50 responses (and many "+1s" written on others' posts, as well as a thread on reddit started by a participant who photographed some of the responses and uploaded it to r/gaming, with over 500 comments [11]. Some of the comments in the reddit thread mentioned what a great question it was to ask people as well, especially children. Compared with the latter question, with only 27 responses (though quite a few +1s), it seemed to be that these more general questions with ethical aspects to them were more engaging than directly asking about solutions. However, the Saturday questions also suggested that balancing a more technical question with a more imaginatively engaging question seemed to encourage more answers to the technical one.

On the Sunday I asked the same question from different perspectives: "What makes you feel good/happy/excited/ accomplished/included in games?" and "What makes you feel sad/uncomfortable/upset/regretful in games?". These got over 50 responses each, with 63+ for the positive question. These questions were actually asking almost the same thing as the first questions on the Friday – but in a more engaging way.

Although I am not doing quantitative research, but qualitative, I feel it is important to look at the engagement at a meta-level – the more engagement the more likely it is to have a good range of responses that cover diverse perspectives. It is possible that the low numbers for the more academic questions was because the answers were "exhausted" and people didn't wish to put down the same answer that was already there. However, I feel this is less likely given the conversations that were had where many people were asking what the more technical questions meant.

On the whole, I found that the question asking was more difficult than I had expected. Coming up with engaging questions that might give me some useful information was very tricky, and involved thinking about what data it was that I wanted from the answers. This allowed me to phrase the questions in more engaging ways as the weekend went on.

2.2.2 Questions at PAX East 2015

The pilot study at PAX East 2014 allowed me to properly engage in 2015 with a full research project behind it. The aim of the research project is to test a portion of the RRI framework set out by the FRRICT project [4, 6]. Although I will not go into the details of that in this paper, as the data has not been analysed yet, I can reflect on the engagement with the questions. Since the post-it note strategy had worked so well the previous year, I decided to do it again, with more tailored questions for my research. Additionally, I had ethics clearance to perform little interviews where the participant either had a response too long for the post-it or where I wanted to know more information about why they'd written what they'd written. This allowed for several much deeper sets of data for use in later analysis. These were recorded using the Apple iPhone built in voice recorder function. 17 interviews were collected in this way across the weekend.

The questions asked over the weekend were as follows, with response numbers in brackets:

- What is a really fun or cool thing about video games? (74)
- What qualities make for a great male character? (38)
- What is really not fun or cool about video games? (64)
- How have video games changed your life? (41)
- What advice might you give to the video game industry (44)
- What qualities make for a great female character? (38)

Compared with the year before, in general there was a much higher engagement with the booth than with the more academic questions of the previous year. Considering there were also the 17 small interviews, despite not having analysed the data from this experience I think there is a much richer set of data available from the 2015 PAX East effort. Anecdotally, I also noted fewer explanations were required about what an ethicist was, and it felt that participants were finding these questions quite easy to respond to. It's possible that there was more incentive to become involved in questions about sexism and gender (and other ethical issues) due to the Gamergate debacle that began in August of 2014 [9] (well after PAX East 2014, and not long before PAX East 2015). Gamergate highlighted a lot of sexist activity in video games, and I was asked several questions about it while staffing the booth. Some of the post-it notes also explicitly mention Gamergate. Additionally, since one of Gamergate's supposed goals was "ethics in video game journalism" (a discussion outside the scope of this paper) the reduction in explanation of what an ethicist is might be explained through greater exposure of video gamers to the concept of ethics. However, even if increased engagement is due to controversies outside of PAX East, this should not impact the quality of the results.

3. RESPONSIBLE RESEARCH AND INNOVATION IN INDUSTRY

Responsible Research and Innovation (RRI) is an increasingly used term that describes possible methods and approaches to ensuring technological innovations are in the interest of society, and for society [2, 10]. While it is still a fairly new concept, and frameworks are still being developed and tested to effectively incorporate RRI ideas into mainstream research and innovation practices, there is still much research to be done in effective engagement at all levels of the process, from policy to industry to civil society organisations, and to end users and other stakeholders.

One of the challenges posed by Responsible Research and Innovation in industry is to productively engage with stakeholders in the early stages of development in order to determine social acceptability of the innovation and to identify and mitigate any social or ethical concerns. These *anticipation*, *engagement*, and *reflection* aspects are the cornerstones of the *AREA* framework, culminating in a responsibility to *act* on the findings of these stages [1]. The FRRIICT project took this one step further, identifying different aspects of RRI that could be affected by these stages, namely *people, product, process* and *purpose* [4]. RRI in industry has extra challenges because, unlike publicly funded research, which can be more easily regulated and require ethical and social assessment prior to funding, it is unregulated in this aspect and so requires a more incentive-driven approach. The Responsible-Industry project specifically identified needs of small-medium enterprises for "help to connect with stakeholders" [8].

The method I described in the previous section could be a method by which SMEs can engage with stakeholders to sound out ideas or ask questions that are easily answerable in "post-it" type responses. Another alternative is to work with an ethicist to run a similar booth in a larger stakeholder-centered event where the results could be disseminated back to the companies featured at the event. These could easily integrate into an AREA/PPPP approach through interrogating the Ps through *engagement* and asking appropriate questions in an accessible manner.

Although the answers to the questions asked at PAX East are not so relevant to this paper, it is important to note that as a data collection tool the combination of the discussion aspect of the booth (for general ideas about how stakeholders feel about things or where their concerns lie), the post-it response questions (which allowed me to gather data about specific questions relevant to my research project), and the short interviews (which allowed for richer data where the post-it response was not clear or where there was not enough room to write it all on the post-it note) allowed me to not only understand further the concerns and feelings of stakeholders about a particular technology (video games) but also to collect those in ways that would allow me to conduct more rigorous research on them. Additionally, this was not just about data collection for my research – attendees genuinely enjoyed discussing questions and asking me about how to deal with ethical dilemmas. While these discussions may not have been directly related to my research, it allowed me to give something back to the community in the form of my expertise, and fostered more engagement because bystanders would often listen in or join in on discussions that were being had, and then contribute to the exercise.

Finally, I have found that it is important how questions to stakeholders are phrased. Questions surrounding particular products (i.e. video games in this case) seem to be the best to get good engagement, as the stakeholders have some experience with it and can understand the questions. Questions about ethics or other theoretical concepts seem to be less accessible, and either become exhausted quickly with possible answers or are not as easy to answer by participants.

4. CONCLUSIONS

Overall this has been a useful exercise – not just in engaging with the community about ethics and responsible innovation from an educational perspective but in understanding more about what stakeholders think about the industry and its products. I would suggest that it could be possible to conduct a similar booth at other industry conventions, such as in medical, robotics, cars, or other events where stakeholders other than industry representatives are presented with industry products in a convention-like setting. ICTs particularly can benefit from this, as conventions and conferences where the general public engage with new and upcoming gadgets and technologies are quite popular. An ethicist running such a booth could provide context and understanding to the industry, translated into RRI-type frameworks to enable companies to develop better technologies. Additionally, for researchers in this area, it can be an excellent data collection tool for research into responsible innovation or other social impact of technology research.

5. REFERENCES

[1] Anticipate, reflect, engage and act (AREA) - EPSRC website: 2015. *https://www.epsrc.ac.uk/research/framework/area/*. Accessed: 2015-06-22.

[2] European Commission and Directorate-General for Research and Innovation 2013. *Options for strengthening responsible research and innovation*. EUR-OP.

[3] Flick, C. 2014. Responsible Innovation in Video Games: Improving Industry Practice. *Proceedings of ETHICOMP 2014* (Paris, France, 2014).

[4] FRRIICT Project 2014. Framework for Responsible Research and Innovation in ICT. *http://responsible-innovation.org.uk/torrii/content/framework*.

[5] Gamasutra - "Diversity Lounge"? PAX has a lot of work to do: 2013. *http://www.gamasutra.com/view/news/207402/Diversity_Lounge_PAX_has_a_lot_of_work_to_do.php*. Accessed: 2015-06-18.

[6] Mittelstadt, B. 2014. *FRRIICT Research Starter Pack*.

[7] PAX East - Boston, MA March 6-8, 2015: 2015. *http://east.paxsite.com/*. Accessed: 2015-06-18.

[8] Responsible-Industry Stakeholder Consultation Workshop Follow-up, Karlsruhe 20 May 2015: 2015. *http://www.responsible-industry.eu/activities/stakeholder_workshop_may_2015/Workshop_Karlsruhe_20150520_Follow-up.pdf*.

[9] Sanghani, R. 2014. Misogyny, death threats and a mob of trolls: Inside the dark world of video games with Zoe Quinn - target of #GamerGate. *http://www.telegraph.co.uk/women/womens-life/11082629/Gamergate-Misogyny-death-threats-and-a-mob-of-angry-trolls-Inside-the-dark-world-of-video-games.html*.

[10] Stahl, B.C., McBride, N., Wakunuma, K. and Flick, C. 2014. The empathic care robot: A prototype of responsible research and innovation. *Technological Forecasting and Social Change*. 84, (May 2014), 74–85.

[11] StrikeAnywherePanda 2014. The best answer for this question. *http://www.reddit.com/r/gaming/comments/22ydho/the_best_answer_for_this_question/*.

Case study research to reflect societal and ethical issues – Introduction of a research implementation plan for ICTs

Emad Yaghmaei
Mads Clausen Institute
Addr. Alsion 2, DK-6400
Sønderborg, Denmark
Tel. +45 6550 9382
Email emad@mci.sdu.dk

Alexander Brem
Mads Clausen Institute
Addr. Alsion 2, DK-6400
Sønderborg, Denmark
Tel. +45 6550 9246
Email brem@mci.sdu.dk

ABSTRACT
The purpose of this paper is to provide the systematic procedures of case study research. A robust case study protocol in the context of societal and ethical issues with a relevant guidelines will set to show how data should collect, present, and analyze. This paper takes in particular the context of Responsible Research and Innovation (RRI) as the phenomenon and investigates on how to apply the methodologies, which could assist us to identify the boundaries between societal and ethical issues and the emerging ICTs. As such, to interpret the collected data and build theory inductively, to have an excellent basis for further qualitative research within RRI context, systematic procedures are needed to conduct. This paper uses a holistic view to the literature for providing guidelines for conducting and reporting case study research for RRI context. It defines the information that needs to be gathered from the cases, the way this data is to be analyzed and the processes of reflections to be undertaken. Checklists for conducting the case study protocol are linked to each step of systematic procedures and applicable for researchers, ICT managers, reviewers, and readers to allow them taking account of societal and ethical aspects, in particular RRI principles in emerging ICTs in a proper way.

Categories and Subject Descriptors
K.4.0 [**Computers and Society**]: General

General Terms
Management

Keywords
Case study, Guidelines, Responsible Research and Innovation, Checklists, Qualitative Research Methods

1. INTRODUCTION
Societies need to think about how humans and their interactions with future technologies should be taken into account. This requires an in-depth understanding of present human interactions with technologies. To achieve understanding of such real-life phenomenon in depth, substantial contextual conditions should be addressed to identify the nature of societal and ethical aspects of technologies. As new technologies spread further into our personal and social lives [11], as the acceptance of emerging technologies and their effects on personal lives is continuously growing, modern societies need to solve current problems of the future faces; despite the uncertainty of the future and difficulty of studies on novelty, one may investigate the recent concept of addressing responsible research and innovation (RRI) principles in emerging technologies. In this spirit, emerging information and communication technologies (ICTs) have a highlighted role of current changes in our lives. As such, considering direct interactions between humans and emerging ICTs as well as finding out how we could integrate RRI principles into research and development phase of projects in emerging ICTs are our main concerns in this work. While societal and ethical aspects in emerging ICTs will light up as the foreseeable future, the current approaches of how address these issues are problematic; thus, we need to identify what kind of method we should apply to investigate such contemporary phenomenon, how, and through which ideal steps?

In seeking to investigate on emerging ICTs for addressing human concerns, a single analytical research method itself might not be able to corroborate the research objectives [31]. To overcome the problems of investigating on such a contemporary phenomenon, the case study methodology is well suited for generating deeper understanding of issues that are hard to investigate in isolation [7, 31]. In essence, studying of a contemporary phenomenon in depth in a real-life context, providing novel theoretical insights, or developing studies in a holistic setting all could be met by applying case study research.

The ability to apply a well-suited case study research is of particular matter given the enhanced need in addressing qualitative research methods in a broad field of study [10]. In spite of criticizing the case study research for lack of rigor, flaw in providing scientific generalization; and being biased by investigators, the critique might be fulfilled by applying suitable case study protocol with a proper guideline for understanding the study contemporary phenomenon in its natural context in depth [8, 12, 22]. In practice, while few researchers are succeeding in achieving proper understanding of societal and ethical aspects in emerging ICTs by applying suitable case study methods, there are many more researchers despite being expert lack the systematic procedures to apply for achieving that success [31]. Thus, the need for having a suitable set of systematic procedures that contribute to the quality of the research is needed in the literature.

Therefore, the final questions picked up by us in the sense that what are the systematic procedures steps –ideal steps- for conducting case study research? And how to apply it in addressing societal and ethical issues in the emerging technologies? These questions critically dig up some interesting action points from an

action project to recommend researchers, ICT managers, reviewers, and readers in stressing societal and ethical aspects in emerging ICTs.

This work renders the main action plans for the case study research method. It explains the scope of different qualitative research methodologies, details approaches, and render the ultimate checklist for case study research steps.

2. BACKGROUND

2.1 Qualitative Research Method

Qualitative research methods would develop a strong applied orientation either for addressing research questions or designing methods procedures [10]. Different types of such research methods would apply in a variety of disciplines such as social sciences, human sciences or socio-technical systems. In essence, qualitative research claims to investigate a complex phenomenon "from the inside out" [10, p. 3] and describe real-life issues in-depth (e.g. interactions between humans and emerging ICTs) for gaining deeper understanding of social realities. As such, in our work, a qualitative research method; in particular case study research is set to investigate RRI through a more involving and open approach rather applied approaches of other research strategies; those strategies are more objective-oriented and may raise large standardized quantities and normative concepts [10]. While for addressing RRI principles in emerging ICTs through standardized methods, one may design his/her fixed data collection instruments (e.g. questionnaire) to gain fixed outputs about the subject of investigation, qualitative research methods scope is set for more being open to what is new about the phenomenon studied. In this way, perceptions of societal and ethical aspects can be described likewise their boundaries with emerging ICTs might be identified in much tighter and more focused way. On this basis, we evaluate the different qualitative research methods and their likely contributions to the field. By so doing we choose case study research as our project research methodology.

Hence, in order to choose an applicable method in our project, the logics of the qualitative research approaches need to be well-described. In this context the concept of case study research is gaining our aim. This paper follows the most widely used definition of case study research, which is that one Yin [31] argues:

"… empirical inquiry that investigates a contemporary phenomenon in depth and within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident." [31, p. 18]

It is worth noting that in order to be able to identify societal and ethical issues in emerging technologies, we would have to set an appropriate scope for our empirical study. Considering this fact that societal and ethical issues comes from fundamental changes to human entails such as human capabilities, choices, privacies, etc. we therefore need to address these societal and ethical changes in emerging ICTs; in doing so, three other major qualitative research methods might also be applied:

- Ethnography, or participant observation, which is a social research method designs as "… it does on a wide range of sources of information. The ethnographer participates, overtly or covertly, in people's daily lives - for an extended period of time, watching what happens, listening to what is said, asking questions; in fact collecting whatever data are available to throw light on the issues with which he or she is concerned". [20, p. 1]
- Action research, which is conducted either for solving a problem or generate new insight and that characterized by researchers' interaction with people involved in the study [4]. More precisely, "Action research may be defined as an emergent inquiry process in which applied behavioral science knowledge is integrated with existing organizational knowledge and applied to solve real organizational problems. It is simultaneously concerned with bringing about change in organizations, in developing self-help competencies in organizational members and in adding to scientific knowledge". [29, p. 439]
- Grounded theory, as the comparative analysis' general methodology is taken as "… a systematic, inductive, and comparative approach for conducting inquiry for the purpose of constructing theory" [1, 2, 3] and that "grounded theory must fit the situation being researched as the categories, must be readily (not forcibly) applicable to, and indicated by the data under study". [16, p. 3; 30, p. 12]

Here, there is a set of facts worth to state. While each single analytical research methodologies primarily set its own purpose of research, either exploratory, explanatory, descriptive, or improving research purposes [28], it is important to note that all those purposes could be fit partly by a certain research method that is case study research.

Reference to characteristics of research methodologies explains key aspects of ethnography such as: being mainly 'descriptive' method to portray a phenomenon and try to tell the story, the emphasis on development and testing the theory [6, 15] and "developing an alternative view of the proper nature of social research" [20, p. 6]. Based on these theoretical aspects, ethnographic methods, like participant observation, unstructured interviews, and archival materials may be applied in case study research.

Likewise, action research method' primary objectives may highlight this methodology mainly as an 'improving' approach of a certain aspect of the subject of investigation. This paper follows the idea of understanding action research as a philosophy of life [27] and, in line with Greenwood [17], sees it as:

"… action and research, reflection and action in an ongoing cycle of cogenerative knowledge". [17, p. 131]

As such, the purpose of generating rigor actionable knowledge from a robust action research fits closely to case study research objectives [28]. In light of these insights, action research methods, like semi-structured interviews, may reflect on data collection procedures of case studies.

In spite of rising contrasts between and within schools - Glaserian school and Straussian school- in using the grounded theory, a common view from ideation to grounded theorizing, in line with Locke [23], will follow in this paper in that grounded theory is:

"…Moving from theory that was developed by thinking things through in a logical manner to theory developed from rich observational data." [23, p. 19]

This perception captures the critical points that reflect in the grounded theory: compatibilities of the grounded theory and symbolic interactionism [14], being mainly 'explanatory' method in addressing conceptual framework of phenomenon studied, and

the modifiability of this theory/methods package [13, 15]. As such, the grounded theory by its methods, like participant observation, semi-structured interviews, and archival materials, by having a symbolic interactionism theoretical background could address fundamental processes through 'why' and 'how' questions.

These insights lead us to a more holistic view on qualitative research method in that either of above-mentioned methods was originally used for a certain research purposes either exploratory, explanatory, descriptive, or improving objectives; although case study methodology as a versatile research method may hold all purposes in all.

2.2 Case Study Research in Emerging Technologies

The nature of research question(s) is probably the most fundamental rationale behind choosing right research method [31]. In essence, research question(s) should be traced over time in dealing with study' operational links. For contemporary research areas where we have relatively little knowledge about the topic, when we miss adequate literature with profound practical experience, investigation on relevant phenomenon within its real-life context in-depth through a concrete research question(s) would be suggested [31]. In these cases, theory building from case study research is a common way to create a theoretical basis using empirical evidence [8, 31]. In essence, in seeking to develop insights about individuals, groups, and organizations, and about processes, relations, and related changes, case studies are commonly used in sociology, political science, social work, and business and technology [31]. It is therefore reasonable to conduct case studies in areas where we are going to investigate humans and their interactions with technologies. Interacting of humans with emerging ICTs lead to a numerous related artifacts in that the most important item needed is increasing knowledge about social and individual lives. Research on addressing societal and ethical aspects of emerging ICTs is to a large extent aimed at studying on how the changes in ICTs and their consequences affect the interaction way of humans with the world. Emerging and enabling ICTs is increasingly borderless and, given the complexity of humans and their interactions with ICTs, it is worth to compare processes, relations, and related changes on the matter, across involved stakeholders - individuals, groups, and organizations. In light of this issue, since both societal and ethical issues –the phenomenon- and the emerging ICTs – the context- are highly pertinent to each other and suitable for case study research, understanding fundamental contextual conditions helps to carry on designing case studies and distinguish this method from the other research methods.

In addition, nevertheless case studies and other qualitative research methods may overlap, the case study unique position is secured as it could be dealt with a full variety of evidence – documents, artifacts, interviews, and observation- beyond what might be investigated in any other single research methods. In fact, while case study research' primary objectives may highlight this methodology mainly as an 'exploratory' approach of the subject of investigation, it might integrate 'descriptive', 'explanatory' and 'improving' perspectives on the phenomenon studied.

Hence, the boundaries between societal and ethical issues –the phenomenon- and emerging ICTs – the context- should be cleared, interpreted and (ideally) inductively connect built theory and collected data. In this spirit, the aim of the researcher should be to build the theory from the generated data by detecting patterns within the findings and afterwards hypothesize, which is consistent with case study research [9]. As such, the case study research covers the logic of design, follows data collection techniques, and conducts specific data analysis approaches to study on a contemporary phenomenon in depth within its real-life context, which all these aspects fit to our study aims when we are seeking to address how to identify the societal and ethical issues related to emerging ICTs.

2.3 Case Study Protocol Steps

The process of conducting a case study protocol in almost any kind of empirical research includes major steps covering different design components; these components provide an overview for the building theory process:

I. Case study design: Research question(s) of the case study in accordance with objectives of study are set and the case study is getting started.

II. Data collection design: instruments and protocols for data collection are crafted, the case selection criteria are set, and unit(s) of analysis is/are specified.

- Field notes: The logic of linking the collected data in addressing the societal and ethical issues of emerging ICTs is identified.

III. Analysis of collected data: After collecting data from each individual case, cross-case pattern would be applied besides using divergent techniques to interpret the findings and analyze data.

IV. Shaping outcomes and enfolding the literature

V. Reporting: highlight the lessons to be learned

This process must have a flexible design strategy in where a numerous iteration steps should apply to reach theoretical saturation [7]. Thus, steps II and III may be conducted incrementally according to set objectives and research question(s). This work does not offer the space to discuss step IV in details independently; therefore its terms pop up at step III and V.

Collectively, during the stage of conducting case study research methodology, the project has to design a robust case study protocol with its relevant guidelines (action plans) to collect, present, and analyze data fairly in domain of ICTs. Therefore, cases are based on the area of ICT. A further focus of case study protocol in this project is to provide the necessary academic rigor to record, analyze and synthesize the comparative cases to assess societal risks and ethical issues. The step of rigorous scientific observation and analysis is essential to demonstrate potential benefits for emerging ICTs at large to follow up societal and ethical issues.

The purposes of the following action plans for addressing the societal and ethical issues related to emerging ICTs are to guide readers in determining when, where, and how case study activities are conducted and who is responsible for those. Hence, items that be considered are: (I) Case study design (II) Data collection (III) Data analysis (IV) Case study reporting

3. CASE STUDY DESIGN

Our case study design contains a set of activities, instruments, procedures, and general rules. In essence, case study design assists us to protect objectivity by providing explicit descriptions of the steps to be taken. Such design contains information on the specific questions/objectives addressed by the study, the description of the

case(s) and its units of analysis, describing contextual events surrounding of cases, the search strategy for identification of relevant studies, and the rationale behind the selection part for inclusion and exclusion of studies in the review [5]. Peer review on case study design will mitigate the risk of missing relevant data sources, optimize interview questions, and address specific roles to include in the research to assure concrete relation between research questions and interview questions. Finally, to address changes during the research project accurately and track them accordingly, the case study design would be monitored regularly to avoid any mistakes.

This paper categorizes key findings and the overall focus of the case study design and will summarize a set of items related to protocol design.

Our study objective is how to how to implement societal and ethical issues along the value chain of ICT. This main research question would be focused along the whole project. In reflecting societal and ethical issues in emerging ICTs, the question raises that what is the most straightforward approaches in selecting case(s) to look for data; the case may be an individual, a team, a project, a product, a technology, an entity, etc. having settled on the design principles of the case(s), the next question is what is the unit of analysis within the case. Likewise, information about the relevant individual, group, project, product, technology, or company could be collected for the unit of analysis.

In principle, the embedded unit(s) of analysis would take within this project. Embedded case studies design, where multiple unit(s) of analysis is studied within a case, is our choice to take for case study protocol.

In essence, this paper is going to look up at four ICT companies as cases while we are seeking to study on four distinctive projects under those companies as units of analysis. All projects, obviously, are in the domain of ICT.

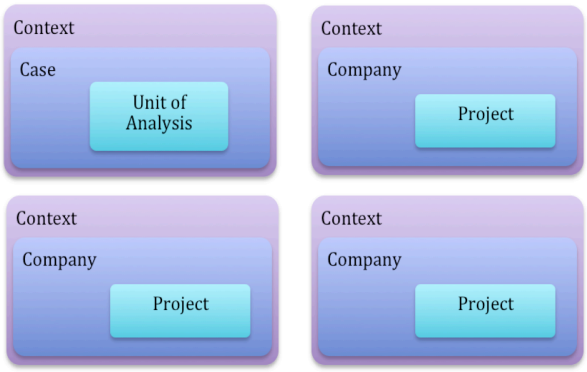


Figure 1. Context: ICT domain
Case: ICT companies
Unit of Analysis: an ICT project inside of cases

Since the eligible number of cases and units of analysis is large, a three-stage screening procedure is taken. The first round of screening process contains collecting archival data from candidates in order for setting the operational criteria. In our example, series of theoretical considerations was undertaken to generate case selection criteria; after through clustering the candidates, the third screening stage applied by which we select four cases, one pilot case, and two alternative cases. Naturally we also revisit our earlier operational criteria along the case selection phase regularly - iteration process - to assign suitable projects as units of analysis.

In order to present our case studies consistent and coherent and having an approach that is scientifically sound, we should to a large extent control what data is collected by defining cases and their relevant units of analysis in terms size, domain, process, subjects, etc. adequately. At case study design phase, the cases could distribute either across a specific industry sector or a range of industry sectors depends on the research question logic. In our example, case studies are designed by interviewing four companies perceived to be actively developing a range of ICTs. Our criteria for the selection of relevant projects were as follows:

- Actively engaging in research and innovation activities for developing ICTs (Hardware-oriented or software-oriented)
- Recognition as reflecting societal risks and ethical principles in firm policies
- Coverage different source of funding for projects (from purely public funded projects to purely private funded ones)

3.1 Quality of Research Design

One common aspect of research design is that how to check case studies with regard to the quality. According to guidelines and tactics in judging the quality of research designs [7, 21, 31], the high quality case studies follow four design tests include:

- Construct validity
- Internal validity
- External validity
- Reliability

The cases and units of analysis must explicitly check in terms of using multiple data sources on data collection, should maintain a chain of evidence to allow reader to follow data from initial research steps to ultimate conclusions in a feedback loop style, and have a rigorous scientific observation step by peer-review to fulfill construct validity. In our case studies, we involved our investigators in collecting evidences from different available sources in that we could triangulate the data. The peer review for the quality assurance purposes is applied at any stage.

3.2 Ethics in Case Study Research

Regarding activities involving human research i.e., the case study research, issues of research ethics must be considered [18, 19]. There is a set of actions that must be fulfilled to gain ethical approval for the case study research, particularly in our work for ICTs cases. Main activities in case study that need to be ethically assessed include:

- Gathering information from or/and about individual human beings (and firms) through:
  - Interviewing
  - Surveying
  - Questionnaires
  - Delphi study
  - Focus group(s)
  - Observation(s) of human behavior
- Using archived records/documentation in which individuals/firms are identifiable.

- Researching into activities which involves direct observations of or contact with those who engaged in case studies as a participant-observation
- Research which involves a technological device, a tool or instrument, or few other physical/cultural artifacts
- Researching into activities that might impact on human behavior e.g. behavioral studies

All involving stakeholders must explicitly consent to participate in case studies. In our work, we captured the experience and opinions of ICT industry thought leaders on the concept of societal risks and ethical issues through interviews, collected quantitative data from all kind of stakeholders (industry, researchers, users and policy makers) by applying Delphi exercise, and obtaining the experience from group of experts in focus groups. Since the handled information is confidential we used a consent form along all activities. A separate form is required for each activity.

Beside a consent form, accompanying participant information letter to support the application is needed to invite participant into the research project.

The participants should be informed about the research outcomes; such set of feedback during research activities enhances the validity and fosters reliability of the research. Moreover, all research activities must be recorded in some ways, as it is not possible to take full notes and conduct an effective action at the same time. In addition, before using transcripts of interviews and observations, researchers must ensure collected data, by removing any reference to particular persons or companies, is anonymised.

At the end, our case study design is adapted to the needs in RRI in the emerging ICTs and applies related terminology and guidelines mostly from social science and information systems.

4. DATA COLLECTION

Different data sources will apply in doing our case studies in order to mitigate the risks of one interpretation of one single data source. In essence, our case study protocol takes into account different viewpoints and roles with regards to project path and uses as many sources as possible. Such multiple sources of evidence allow our investigations to address broader range of relevant issues, converges data in a triangulating fashion, and develops lines of inquiries. We are seeking, therefore, to use an impressive array of sources of evidence and extend our observations, with semi-structured interviews, archival records, and midstream modulation method.

In the main scenario, this paper tended to conduct fully transcribed semi-structured interviews with five project managers and members in two rounds, apply two rounds of collecting project documentation and archival records (product flyers/financial reports/archival records), and circulate survey data among 10 end users of products. In principle, To optimize our observation upon the company's activities in the area of RRI, the understanding end users' feedbacks to product(s)/service(s) will be are largely helpful. This paper is seeking to present and observe firm' interaction with end users. In doing so, we invited users to fill in a survey concerning their experience with the project / product. The authors will work with the company in developing, distributing and analyzing the survey.

Alternatively, however, it was supposed to conduct "Midstream Modulation (MM)" as qualitative method and "WIAT+" as quantitative method to measure the impact of societal and ethical issues on industrial projects. This method, however, did not choose for the final data collection due to lack of having practical experience on acting professional human scientists. In fact in this method, our investigators were supposed to go to study on four assigned managers/members of each selected project for twelve subsequent weeks (2 assigned project members play MM group and 2 remained participants served as comparison group (C-group)). All four project members should have had a background in the domain of ICT, work on the same project but studied different technological aspects of the project.

One of common sources in both scenarios is the semi-structured interview. Interviews offer the added advantage of having lines of inquiry of interviewer, simultaneously asking controversial questions in a focused, or in-depth manner. In fully structured interviews, however, researcher is looking for a meaningful causality between constructs; such interviews may be seen as a formal survey, have been standardized, and conduct with closed questions.

In contrast, interviews could be held in an absolute unstructured manner in which interviewer guide conversations rather than being fixed with questions. Such interviews may therefore prolong over an extended period of time. In essence, these interviews with highest rate of flexibility use likely for oral or life history interviews, group interviews such as focus group, and survey interviews [25, 26]. Unstructured interviews are set to explore on how people qualitatively experience the phenomenon.

To explore and describe of the societal and ethical issues of the emerging ICTs, the mix of open and closed questions is recommended to see individuals' interactions with the phenomenon qualitatively and quantitatively. Hence, semi-structured interview is applied in our case studies. The initial findings of the pilot case were coded, categorized, and evaluated to allow us to edit interview guideline and recapitulate what the emerging ICTs should represent responsibly and ethically.

In our cases, we notify interviewees with regards to project ethical approval and ensure them that gained information from interviews, sensitive scientific and technological details of projects will only be used for the our project and will not be used for any other purpose.

In a nutshell, for the current case study method, we visit the cases on specific occasions over a time span of twelve weeks. During these visits we conduct interviews with approximately five members of staff. We do interview each member two or three times.

Personal field interviews are held with the CEO/high level strategy manager, the CTO/high level tech manager, the marketing or sales or CSR manager, the R&D manager, and a member of R&D staff. In doing so, we send interview participant information sheet and a consent form to all potential respondents in which tell them it is important that they understand why the research is being undertaken and what their participation will involve. Moreover, to permit a triangulation of data and provide valid observations, we review the company records, in particular project records. Hence, we ask the companies to provide project materials with regard to sustainability, responsibility, codes of conduct, ethical reports, corporate social responsibility (CSR) reports and etc. We assure them that we will treat any such records as confidential and are prepared to sign relevant non-disclosure agreements.

To optimize the study' result in the context of RRI, understanding end users' feedback to product(s)/service(s) will be helpful. We

would tend to observe the company interaction with end users. As such, we invite users to fill in a survey concerning their experience with the project / product. To do so, we work directly with the case companies in developing, distributing and analyzing the survey.

5. DATA ANALYSIS

Qualitative data analysis methods are applied for our case studies. Such analysis contains examining, categorizing, tabulating, and testing to draw empirically based conclusion [31]. The basic aim of the project analysis, which is obtaining conclusions from the data, induced by keeping a clear chain of evidence. We carry out case analysis in parallel with the data collection processes. In addition, since our case study follows a feedback loop, new insights during the analysis may require us to redesign the original research design. To investigate such new insights and discoveries, we update interview guidelines to gain new data in controversial cases. To reduce biases of analyzing, at initial steps of data analysis peer-reviews assist us to optimize the authenticity of the process. In fact, the preliminary results from data collection are transferred into a common analysis process and invariably we keep tracking and do report at analysis stage to increase the validity of the study.

This study applies NVivo as computer-assisted qualitative data analysis software. The tool assists us to code and categorize large amount of collected data as narrative text that conducted by semi-structured interviews, MM, and survey data. NVivo also codes and categorizes large volumes of written materials, such as archival records, press articles, etc. NVivo is compatible with SurveyMonkey as data collection tool. Therefore, we use those for both collecting data and data analysis.

The analysis of qualitative data is conducted in a series of steps [24]. First data is coded, which means we put information into different arrays in order to code the text according to certain theme, area, construct, etc. we also add investigators' comments into the coded data (with codes or sub-codes), such as memos. As its result, we make a matrix of categories and locate the evidence within those categories. To examine the data, we create data displays, includes flowcharts and other graphics. Such flowcharts help us to identify a first set of hypotheses. Primary hypotheses could be frequent phrases in different parts of material. These identified hypotheses, which use in parallel with data collection under an iterative process, would be optimized regularly; in turn a small set of generalization can be formulated. These series of steps are executed iteratively to meet the ultimate goal.

Our analysis will show that we utilize all the collected sources of evidence. We develop rival hypotheses, cover key research questions, and use as much evidence as available. Likewise, our interpretations would account for all the sources of evidence and our analysis sought to address as many rival interpretations as possible. We address irresponsible research and innovation along the value chain of ICT and seek to reflect both responsible and irresponsible innovation aspects into our analysis and interpretations.

Finally, most significant aspects of case studies would address to reflect our expertise in carrying out the analysis since one may need to have a careful and detailed analysis work in this section.

6. REPORTING

The case study report tells readers what the study is clearly about; explicitly the case study elaborates the societal and ethical issues related to the emerging ICTs, shows itself as a significant communication device to communicate a clear sense of the studied case, provide a history of the inquiry, indicates basic data in focused form, and articulate robust conclusions for readers.

In essence, the multiple case studies consist both the single case studies and some cross-case chapters. In our case study, first we coverage single issues such as presenting CEOs perceptions about RRI principles as separate chapter and then conduct the cross-case studies, which appear at the end of report in that the societal and ethical issues related to the emerging ICTs would monitor through the lens of all industrial stakeholders.

Our case study reports are regarded as part of a larger, mixed method studies that other methodologies would include.

To increase construct validity of the case study report one may argue to have a review procedure in which peer-reviewer comments assist to rewrite and optimize the report.

7. SUMMARY

The case study research is conducted for corroborating substantial understanding toward the investigation of contemporary phenomenon. Case study method is a suitable methodology for social science research. Similar to most case study research on social science, this study was focused on framing of describing, understanding of the unclear boundaries between RRI as the phenomenon and the emerging ICTs as the context in real life.

This work aims to provide a concrete checklist for conducting and reporting case study research to reflect societal and ethical issues in the emerging ICTs. For having such as checklist, a set of iteration stages in design phase should set out. Research question in conjunction with research objectives is first determined. Instruments and protocols for data collection are crafted, the case selection criteria are set, and unit(s) of analysis is/are specified. For our certain case study, the logic of the linkage of collected data with the societal and ethical issues in emerging ICTs is identified. The data collection methods are namely interview, observation, archival records, etc. After collecting data from each individual case, cross-case pattern would be applied besides using divergent techniques to interpret the findings and analyze data. Finally, the researcher needs to shape outcomes and enfolding the literature, highlight the lessons to be learned, and report sufficient data.

Similar to other guidelines this checklist is also need to assess in practical scene to corroborate all its action points. Further results will come after checklist evaluation while this paper has set relevant checklist that can be developed to further studies, in particular in reflecting societal and ethical issues in emerging ICT.

8. REFERENCES

[1] Bryant, A., and Charmaz, K. 2007. Introduction - Grounded theory research methods: Methods and practices. *The Sage Handbook of Grounded Theory.* edited by Bryant, Antony; Charmaz, Kathy, Sage, London, 2-28.

[2] Charmaz, K. 2006. Constructing grounded theory: A practical guide through qualitative analysis. *Sage, London.*

[3] Charmaz, K., and Henwood, K. 2007. Grounded Theory. In: C. Willig and W. Stainton-Rogers Ed. *The SAGE Handbook of Qualitative Research in Psychology*. SAGE, London, Publications Ltd. 240-260.

[4] Coghlan, D. 2011. Action Research: Exploring Perspectives on a Philosophy of Practical Knowing. *The Academy of*

Management Annals, 5, 1, 53-87. DOI=10.1080/19416520.2011.571520.

[5] Davies, H. T. O., and Crombie, I. K. 2000. Bias in case-control studies. *The British Journal of Hospital Medicine,* 61, 4, 279-281.

[6] Denzin, N. K. 1978. The research act: A theoretical introduction to sociological methods (2nd ed.). *New York: McGraw-Hill.*

[7] Eisenhardt, K. M. 1989. Building theories from case study research, *Academy of Management Review,* 14, 4, S. 532–550.

[8] Eisenhardt, K. M., and Graebner, M. E. 2007. Theory building from cases: Opportunities and challenges. *Academy of management journal,* 50, 1, 25–32.

[9] Farquhar, J. D. 2012. Case Study Research for Business. *Sage Publications Ltd.*

[10] Flick, U., Von Kardorff, E., and Steinke, I. 2004. What is Qualitative Research? An Introduction to the Field, in: A Companion to Qualitative Research, edited by Uwe Flick et al, *Sage, London,* 3-11.

[11] Floridi, L. 2007. A look into the future impact of ICT on our lives. *The Information Society,* 23 1, 59-64.

[12] Flyvbjerg, B. 2007. Five misunderstandings about case-study research. *In Qualitative Research Practice: Concise Paperback Edition.* Sage, 390–404.

[13] Glaser, B. G. 1978. Theoretical sensitivity: Advances in the methodology of grounded theory. *Sociology Pr.*

[14] Glaser, B. G. 2005. The grounded theory perspective III: Theoretical coding. *Sociology Press.*

[15] Glaser, B., and Strauss, A. 1967. The Discoveryof Grounded Theory. *Aldine Publishing Company.* Hawthorne, NY.

[16] Glaser, B., and Strauss, A. 1999. The discovery of grounded theory: Strategies for qualitative research. de Gruyter, New York.

[17] Greenwood, D. 2007. Pragmatic action research. *International Journal of Action Research,* 3, 1/2, 131–148.

[18] Goode, E. 1996. The ethics of deception in social research: a case study. *Qualitative Sociology,* 19, 11-33.

[19] Guillemin, M., and Gillam, L. 2004. Ethics, reflexivity, and "Ethically Important Moments" in research. *Qualitative Inquiry,* 10, 2, 261-280.

[20] Hammersley, M., and Atkinson, P. 2007. Ethnography: Principles in Practice. *Taylor & Francis. ISBN:* 0203944763, 9780203944769.

[21] Hoon, C. 2013. Meta-Synthesis of Qualitative Case Studies: An Approach to Theory Building. *Organizational Research Methods,* 16, 4, 522-556.

[22] Lee, A. S. 1989. A scientific methodology for MIS case studies. *MIS Q,* 13, 1, 33–54. DOI= 10.2307/248698.

[23] Locke, K. 2002. Grounded Theory in Management Research. *Thousand Oaks: Sage.*

[24] Miles, M. B., and Huberman, A. M. 1994. Qualitative data analysis: An expanded sourcebook. *Thousand Oaks. CA: Sage.*

[25] Minichiello, V., Aroni, R., Timewell, E., and Alexander, L. 1990. In-depth Interviewing: Researching people. *Hong Kong: Longman Cheshire Pty Limited.*

[26] Punch, K.F. 1998. Introduction to Social Research: Quantitative and Qualitative Approaches. *Thousand Oaks: Sage.*

[27] Reason, P., and Bradbury, H. 2008. *Handbook of action research* (2nd ed.). London: Sage.

[28] Robson, C. 2002. Real World Research. Blackwell, (2nd ed.).

[29] Shani, A.B. (Rami), and Pasmore, W.A. 1985. Organization inquiry: Towards a new model of the action research process. In D.D. Warrick (Ed.), Contemporary organization development: Current thinking and applications (pp. 438–448). Glenview, IL: Scott Foresman [Reproduced in D Coghlan & A. B. (Rami) Shani (Ed.). (2010). *Fundamentals of organization development*, London: Sage, 1, 249–260.

[30] Strauss, A., and Corbin, J. 1998. Basics of qualitative research: Techniques and procedures for developing grounded theory, 2nd Ed., *Sage Publications: Thousand Oaks.*

[31] Yin, R. K. 2009. Case Study Research: Design and Methods, fourth edition, *Thousand Oaks,* CA: Sage Publication.

A Realisation of Ethical Concerns with Smartphone Personal Health Monitoring Apps

Tilimbe Jiya
De Montfort University,
The Gateway, Leicester
LE1 9BH
Telephone +44 01162507475
Email:tilimbe.jiya@email.dmu.ac.uk

ABSTRACT
The pervasiveness of smartphones has facilitated a new way in which owners of devices can monitor their health using applications (apps) that are installed on their smartphones. Smartphone personal health monitoring (SPHM) collects and stores health related data of the user either locally or in a third party storing mechanism. They are also capable of giving feedback to the user of the app in response to conditions that are provided to the app therefore empowering the user to actively make decisions to adjust their lifestyle.

Regardless of the benefits that this new innovative technology offers to its users, there are some ethical concerns to the user of SPHM apps. These ethical concerns are in some way connected to the features of SPHM apps. From a literature survey, this paper attempts to recognize ethical issues with personal health monitoring apps on smartphones, viewed in light of general ethics of ubiquitous computing. The paper argues that there are ethical concerns with the use of SPHM apps regardless of the benefits that the technology offers to users due to SPHM apps' ubiquity leaving them open to known and emerging ethical concerns. The paper then propose a need for further empirical research to validate the claim.

Categories and Subject Descriptors
K.4.1 **[Computers and Society]:** Public Policy Issues – *Ethics, Privacy, regulation*

General Terms
Human Factors

Keywords
Ethical Concerns, Smartphone Apps, Personal Health Monitoring, Ethics, Mobile Health Applications

1. INTRODUCTION
As smartphone health monitoring applications (SPHM apps) move from an introductory stage through societal permeation to a stage where they are widely used called the power stage [17] there are ethical concerns that needs to be made aware to users and potential users of these apps. Due to the features of SPHM apps, it is inevitable that ethical concerns will arise and that users and ethicists are aware of these. The knowledge or realisation of these ethical concerns is very important not just to the user of SPHM apps but to the developers of such technology and other stakeholders. One of the reasons for its importance is that it promotes responsible innovation through a feedback mechanism that virtually exist between the app users and developers on how to address them, either in existing or future products.

Thus said, this paper will attempt to answer the question on what are the ethical concerns with personal health monitoring apps on smartphones, viewed in light of general ethics of ubiquitous computing. Some doubt the relevance of ethics in computing technology [20] however, this paper argues that it is reasonable to suppose that the growing ubiquitous nature of SPHM apps necessitates reflection of related ethical concerns in society and calls for a stronger awareness of computer ethics in society.

The realisation of these ethical concerns could come about in two ways; firstly by looking at ethical concerns which are coherent with the features of SPHM apps for example its ubiquity and then possibly identifying generic ethical concerns [5] that are likely to be relevant to SPHM apps or secondly we could be more speculative and forecast about potential ethical concerns. This paper will however adopt the first approach, that of identifying generic ethical concerns in light of the ubiquity of the apps. In doing so, the paper conducts a literature survey on generic ethical concerns with smartphone apps and similar technology that fall within the ubiquitous computing (ubicomp) umbrella precisely health monitoring ones.

In this paper two approaches to ethical realisation are used. The first approach is the generic approach from which insights emerge that inform some assumption of ethical concerns and feed into a forecast of some issues that are missed within literature but could potentially arise from the use SPHM apps. This then lays the basis for second approach which is speculative. These speculative

ethical concerns form the foundation of the discussion on the realised ethical concerns of SPHM apps.

The paper attempts to identify ethical concerns of SPHM apps at 3 different levels of the technology in relation to specific features that are relevant to each level of the technology and then, the paper discuses some potential novel ethical concerns of using SPHM apps and propose a need for further empirical research to validate the claim.

2. METHOD

This paper surveyed peer reviewed journal articles and conference papers from three databases namely Scopus, EBSCO Host and Google Scholar focussing on ethics of ubiquitous computing with a keen interest on mobile health monitoring. The search was limited to 10 years due to the novelty of smartphones. However, there was limited literature that specifically used the term *'smartphone'* and more specifically *'smartphone app'* therefore an inference from ubicomp was adopted merely because smartphone apps are part of ubicomp.

2.1 Process

Literature sources were systematically searched within the above mentioned databases. This involved using multiple word combinations and similar words in order to come up with comprehensive search results that were pertinent to inform the paper. Such synonymous words and word combinations as 'personal health monitoring', 'ethical concerns', 'ubiquitous mobile applications', 'pervasive applications', 'smartphone health apps' etc. were used in the literature search. The search results were scrutinised to ensure that the technology discussed was closely related to smartphone health monitoring apps according to the features described in this paper.

2.2 Analysis

As part of the realisation, the paper looked at the generic ethical concerns from articles that were deemed relevant. In Nvivo [3] the content of the paper was analysed in order to develop an overview of the general discourse on ethical concerns that are frequently discussed within the text. Using a word tree and frequency analysis themes were extracted and built across multiple literature sources.

Text is highlighted that had a reference or appeared to make reference to an ethical issue with ubiquitous mobile computing and in some cases smartphone applications. The highlighted text was coded and the coded text segments were assigned themes. From the coding process a discussion of results emerged based on the frequency of code appearance. 5 themes were developed referring to the main generic ethical concerns realised from literature.

3. RESULTS

The literature survey resulted in 87 results and using the inclusion criteria only 27 were relevant from which 16 were selected as the most apt after reading them. The inclusion criteria used in the paper was; the age of articles had to be not more than 10 years; the main focus of the papers should be related to ubiquitous mobile applications that are related to health monitoring and their outcome should potentially discuss or suggest ethical concerns or issues with technology.

From the selected papers the following generic ethical concerns were found and were categorised into 5 themes that were more recurrent and generally applicable to SPHM apps. These themes are further discussed in the sections below.

3.1.1 Data misappropriation

One of the ethical concerns with using SPHM apps was relating to data misappropriation. The concern arises from questioning the originality of the apps and clarity on where accumulated data is stored and manipulated. Data misappropriation could be defined as the unauthorized use of user's data, without their permission and consent that has potential to result into harm. This is a great responsibility concern.

SPHM apps are developed by developers who are both regulated and un-regulated. Data can potentially be stored in servers or other storage mechanisms that are prone to malicious compromise either intentionally or unintentionally which could result into harm. A particular concern seem to be the possibility that data could be sold to private corporations and exploited for profit rather than for the public good [19].

As part of personal health monitoring, SPHM apps store personal identifiable data (PID) that could be linked to personal health data (PHD). The combination of the two can be used to identify personal information of the user [18]. Relatively, SPHM apps have the ability to reason with the raw sensor data to identify higher level information, based on established medical knowledge that is embedded within the app [4] and this raw data if fallen in the wrong hands could be used for inappropriate activities that could damage or harm the owner of such data

3.1.2 Identity theft

Another concern that emerged from the literature survey is that of identity theft. This concern is somehow related to the one above merely because identity theft occurs when a user's personal information is stolen and misappropriated to impersonate them for fraudulent activities. Data collected by apps could be used, with a few parameters, to trace even anonymised data back to the data subject in light of re-identification [21] which could then be used for identity theft or identity fraud.

A combination of user's name with other metadata, such as age and location, can identify a user by triangulation [7, 19] and then the user could be impersonated by a fraudster to carry out for instance financial transactions without their knowledge. All SPHM apps especially those that are freely downloaded may share non-personal data on usage which could potentially be combined with the universal device ID or a unique ID of the downloaded SPHM app which could then enable the non-personal data be tracked back to the user therefore identifying them [1, 19]. As mentioned earlier, this is a potential ethical concern because the data has a potential to be used for other unintended activities using the users identified ID.

3.1.3 Privacy infringement

The third concern that frequently appeared in literature is that of privacy infringement. Considering that smartphones are part of ubiquitous and pervasive computing, privacy appears in literature as one of the ethical concerns of mobile apps[10, 16]. The privacy that is discussed here is the one which mostly refers to the separation of user data and personal privacy. This privacy has a direct relationship with security of user data[18], for example during transmissions data could end up in the wrong hands [2, 6]. When people download SPHM apps their privacy is put at risk due to the apps being susceptible to outside invasion thereby affecting the users' privacy. One way that this happens is through SPHM apps that encourage users to share what could be considered sensitive and private information via social media. This is common with activity tracker apps that have their own virtual forums linked to social media in the name of bringing people together for encouragement and sharing of experiences [23].

Privacy infringement in SPHM apps is a resultant of poor data security measures that are put in place within the apps or their features. Many SPHM apps have poor data security due to the way they transmit data[18]. Some SPHM apps transmits unencrypted data using unsecured networks which could be viewed by anyone who is watching or listening on the network [1, 8].

3.1.4 Uberveillance

Another concern that emerged from the literature survey was that of uberveillance. Uberveillance involves identity and location tracking that is constant and embedded in a technology artifact which is real time ad automatic [13] Activity trackers used in SPHM apps can store information about the location and places where the user has been over a certain period therefore leaving a traceable pattern that can be used for uberveillance [12]. Smartphones on which these apps are mostly installed are constantly online and location enabled and rarely do people turn the geo-location-features off when they are out and about [1], as a result they could potentially provide location data which poses a challenge for anonymity for users of SPHM apps.

3.1.5 Legal inadequacy

The last theme that emerge from the literature survey is a concern with legal inadequacy when it comes to SPHM apps. There is lack of policies that govern emerging technologies such as apps and even if policies are in place there is inadequate policing of these policies that guarantees their effective implementation [12, 22]. In addition, the mobile apps ecosystem is unregulated especially with health and fitness monitoring apps and the data that these apps collect is mostly not covered by existing regulations that protect the privacy and security of the personal health information (PHI) [1]. This lack of legal provisions such as privacy protection could have ethical consequences to users such as identity theft and sale of identifiable data by unregulated app developers.

Another point that is of interest is the extent of legal and cultural differences over privacy and other ethical concerns with mobile health apps between global regions, for instance over what constitutes as a medical app and issues around user consent [22][21]. Depending on the resident country of the SPHM app development, both users and developers can be subjected to different laws and legal obligations. Some regions have a weak adherence to the rule of law and limited privacy protection than others therefore users are vulnerable to abuse.

4. DISCUSSION

The pervasiveness of smartphones has facilitated a new way in which owners of devices can monitor their health using applications (apps) that are installed on their smartphones. Personal health monitoring involves behaviour interventions that will promote people's health in reaction to feedback they are receiving from their body or environment [16]. The advancement of mobile technology especially smartphones and ever growing app market, has enhanced mobile personal health monitoring [4, 11, 15]. There has been an increase in the development of apps that can be used to monitor personal health regardless of platform and expertise of user [9] and these have shifted the paradigm of self-health monitoring allowing people to accurately monitor themselves with mobile technology [14].

Regardless of the benefits that this new innovative technology offers to its users, literature shows that there are some ethical concerns to the user of SPHM apps. These ethical concerns are in some way connected to the features of SPHM apps. Smartphones are accessible by society members who have either significant or limited technical knowledge which renders them susceptible to ethical consequences that can result from use of such new technology, in this case, SPHM apps that are available on the consumer market.

During the survey, this paper could not identify any literature that specifically address ethical concerns with smartphone apps, especially those that monitor health. However, this paper managed to find a few that were indirectly related to SPHM apps in respect of its ubiquity. Therefore, this gave the paper a starting point to discuss ethical concerns with SPHM apps.

SPHM apps are built-in, free and/or pay to download from app stores and they demonstrate versatility, usability and functionality at nominal or no cost. Their features which generally include their ability to collect and store health related data of the user either locally or in a third party storing mechanism, render them prone to ethical concerns as a result of loopholes within their functionality for example data being intercepted during transmission. Another feature common with SPHM apps is their geo-location capability which can be used to locate and identify the user. This feature mainly mostly works with the online connectivity of the SPHM app therefore facilitates the online connectivity a real time identification and tracking of the user. As established from the literature survey, this has potential ethical implications to its user. Users of smartphones need to have knowledge on how the 'location' feature works and what sort of information could be sent out merely by not disabling the feature.

SPHM apps are also capable of giving feedback to the user of the app in response to conditions that are provided to the app

therefore empowering the user to actively make decisions to adjust their lifestyle. This is potentially another area of concern because this could result in the user using or misusing of health signals or feedback that they are receiving from their body via the SPHM app. In such circumstances, there is a risk of the user not understanding or inappropriately understanding these signals and leaving themselves prone to risk of self-diagnosing and medicating in attempt to quickly respond to a warning or feedback that they are receiving from their SPHM app. A practical scenario could be a user buying weight loss medication outside their doctor's knowledge, say online, which could potentially result in drug misuse.

From the literature survey we can envisage ethical concerns that are related with SPHM apps from the generic themes that appear from it. Although not a lot is directly pointing at SPHM apps, the concerns discussed in the literature gives us a foundation to speculate more on ethical concerns. In an attempt to speculate on them, this paper proposes a speculative analysis of SPHM apps. This speculative analysis of SPHM apps comprises of 3 levels of the smartphone app technology as shown in Figure 1 below. The figure shows the speculative ethical concerns with SPHM based on a focus at;

i. The features of the app such as its memory capacity that paves way for ethical concerns such as data loss or privacy violation.
ii. The specific artefact and procedures that smartphone apps are involved with therefore looking at different uses and speculative ethical concerns at that level.
iii. At the specific technology i.e. SPHM apps. At this level the speculative ethical concerns are narrowed down to the specific app looking at the specific users, context and features of the app. With regards to SPHM apps, we look at the application or use that happens within certain contexts.

Table 1. A three level speculative analysis of SPHM apps

Technology level - Smartphones		
Focus is on general features of smartphones	Core features • Ubiquity • Sensing • Memory • Invisibility	Ethical concerns • Data loss • Uberveillance • Privacy
Artifact Level – Smartphone apps		
Focus is on specific artifact and procedures	Different uses • Health • Navigation • Temperature	Ethical concerns • Data loss • Data security • Storage issues • Legal inadequacy • Learnability • Privacy
Application level – SPHM apps		
Focus is on specific users and use or context	Application • Mobile health monitoring • Home use • General public • Available on consumer market	Ethical concerns • Misleading health data • Misuses of drugs • Confusion

The first level is the more generic one that looks at smartphone apps and / or ubicomp and then identifies the ethical concerns. This focuses on the technology at large looking at the features that make up the technology i.e. smartphone technology.

The second one is the artifact level where the useful combination of smartphone and other novel technology procedures provides a service or new product. In this case, the consideration is on the combination of smartphone technology and PHM technology to provide a software app that can be used to monitor personal health outside a clinical set up and on the go regardless of users' expertise. The question then becomes are there moral issues that could be presented by this combination of processes and procedures? An example here could now be smartphone apps that store and provide location data that could be used for uberveillance and other unwarranted purposes therefore posing an ethical concern. The combination of the smartphone and the app have a potential of using both features of each different level of the technology therefore represent novel ethical concerns. This shows that as more artifacts emerge, new ethical concerns will be realised.

The third level is the application of SPHM apps. The focus at this level is what context is a SPHM app being used, where is the app being used and who is using it (user characteristics) in relation to the inherent features of the app. Is it for home or professional use? The context in which the SPHM app is used will pose different ethical concerns. An example is when an SPHM app is used by people in 2 different cultural systems the ethical concerns that may arise could potentially be different. In one, the dissemination of app data could not pose as many consequences as in another due to differences in strength and establishment of the regulatory system of the country of origin for the app an what is culturally acceptable or not.

Similarly, the aim of SPHM app would potentially determine the ethical concerns that its use is likely to present. In this case an ideological scenario could be apps that are used for activity monitoring whereby their users are subscribed to a social media group to get tips and offers on products that are tailor-made according to the data provided by the user, will have different

potential ethical concerns to apps that are used for measuring glucose levels in order to prompt the user to take remedial action such as an insulin injection without passing on information to a third party at that particular moment.

5. CONCLUSION

The literature survey shows that there is limited literature that is specifically directed at ethical concerns that affect SPHM apps, however if these apps are considered in the context of their features, an inference of ethical concerns with similar ubiquitous computing devices could be used to realise ethical concerns that affect SPHM apps. With this in mind, ethical concerns of SPHM apps could be realised through speculation on what sort of ethical concerns could emerge at different levels of the technology's focus. Using both literature and speculation of ethical concerns such data misappropriation, identity theft, privacy infringement, uberveillance, legal inadequacy, misleading health data, confusion and potential drug misuse were realised by inferring to ubiquitous computing and multi-level speculative analysis. This highlights a need for more research that is specific to SPHM apps and probably an empirical study of what different stakeholders of the technology think are the existing and potential ethical concerns with SPHM apps.

6. ACKNOWLEDGMENTS

Thanks to Prof. Bernd Stahl and Dr. Catherine Flick for your support towards this paper.

7. REFERENCES

[1] Ackerman, L. 2013. *Mobile Health and Fitness Applications and Information Privacy*. California Consumer Protection Foundation. Privacy Clearing House.

[2] Al Ameen, M. et al. 2012. Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications. *Journal of Medical Systems*. 36, 1 (Feb. 2012), 93–101.

[3] Bazeley, P. and Jackson, K. 2013. *Qualitative data analysis with NVivo*. SAGE Publications Ltd.

[4] Boulos, M.N.K. et al. 2011. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. *BioMedical Engineering OnLine*. 10, (Apr. 2011), 24.

[5] Brey, P.A.E. 2012. Anticipating ethical issues in emerging IT. *Ethics and Information Technology*. 14, 4 (Dec. 2012), 305–317.

[6] Enck, W. 2011. Defending users against smartphone apps: Techniques and future directions. *Information Systems Security*. Springer. 49–70.

[7] Gasson, M.N. et al. 2011. Normality Mining: Privacy Implications of Behavioral Profiles Drawn From GPS Enabled Mobile Phones. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*. 41, 2 (Mar. 2011), 251–261.

[8] Giota, K.G. and Kleftaras, G. 2014. Mental Health Apps: Innovations, Risks and Ethical Considerations. *E-Health Telecommunication Systems and Networks*. 03, 03 (2014), 19–23.

[9] Hayes, D.F. et al. 2014. Personalized medicine: risk prediction, targeted therapies and mobile health technology. *BMC medicine*. 12, 1 (2014), 37.

[10] HowSmartphonesChangingHealthCare.pdf: 2010. *http://www.chcf.org/~/media/MEDIA%20LIBRARY%20Files/PDF/H/PDF%20HowSmartphonesChangingHealthCare.pdf*. Accessed: 2014-10-20.

[11] Istepanian, R.S.H. et al. 2004. Guest Editorial Introduction to the Special Section on M-Health: Beyond Seamless Mobility and Global Wireless Health-Care Connectivity. *IEEE Transactions on Information Technology in Biomedicine*. 8, 4 (Dec. 2004), 405–414.

[12] Michael, K. and Clarke, R. 2013. Location and tracking of mobile devices: Überveillance stalks the streets. *Computer Law & Security Review*. 29, 3 (Jun. 2013), 216–228.

[13] Michael, M. and Michael, K. 2010. Toward a State of uberveillance. *IEEE Technology and Society Magazine*. 29, 2 (2010), 9–16.

[14] Milani, P. et al. 2014. Mobile Smartphone Applications for Body Position Measurement in Rehabilitation: A Review of Goniometric Tools. *PM&R*. (May 2014).

[15] Milosevic, M. et al. 2011. Applications of Smartphones for Ubiquitous Health Monitoringand Wellbeing Management. *Journal of Information Technology and Applications*. 1, 1 (2011), 7–15.

[16] Mittelstadt, B. et al. 2014. The Ethical Implications of Personal Health Monitoring. *International Journal of Technoethics (IJT)*. 5, 2 (2014), 37–60.

[17] Moor, J.H. 2005. Why We Need Better Ethics for Emerging Technologies. *Ethics and Information Technology*. 7, 3 (Sep. 2005), 111–119.

[18] Orwat, C. et al. 2008. Towards pervasive computing in health care – A literature review. *BMC Medical Informatics and Decision Making*. 8, 1 (Jun. 2008), 26.

[19] Rose, N. 2014. The Human Brain Project: Social and Ethical Challenges. *Neuron*. 82, 6 (Jun. 2014), 1212–1215.

[20] Stahl, B.C. et al. 2014. From computer ethics to responsible research and innovation in ICT. *Information & Management*. 51, 6 (Sep. 2014), 810–818.

[21] Tene, O. and Polonetsky, J. 2013. Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Intellectual Property*. 11, 5 (2013), 239–273.

[22] Thomas, C.M. et al. 2013. Smartphones and computer tablets: Friend or foe? *Journal of Nursing Education and Practice*. 4, 2 (Dec. 2013).

[23] Tran, J. et al. 2012. Smartphone-based glucose monitors and applications in the management of diabetes: an overview of 10 salient "apps" and a novel smartphone-connected blood glucose monitor. *Clinical Diabetes*. 30, 4 (2012), 173–178.

[24] VodafoneGlobalEnterprise-mHealth-Insights-Guide-Evaluating-mHealth-Adoption-Privacy-and-Regulation.pdf: *http://mhealthregulatorycoalition.org/wp-content/uploads/2013/01/VodafoneGlobalEnterprise-mHealth-Insights-Guide-Evaluating-mHealth-Adoption-Privacy-and-Regulation.pdf*. Accessed: 2015-01-07.

Animating the ethical demand – exploring user dispositions in industry innovation cases through animation-based sketching

Peter Vistisen
Aalborg University
vistisen@hum.aau.dk

Thessa Jensen
Aalborg University
thessa@hum.aau.dk

Søren Bolvig Poulsen
Aalborg University
bolvig@hum.aau.dk

ABSTRACT
This paper addresses the challenge of attaining ethical user stances during the design process of products and services and proposes animation-based sketching as a design method, which supports elaborating and examining different ethical stances towards the user. The discussion is qualified by an empirical study of Responsible Research and Innovation (RRI) in a Triple Helix constellation. Using a three-week long innovation workshop, U-CrAc, involving 16 Danish companies and organisations and 142 students as empirical data, we discuss how animation-based sketching can explore not yet existing user dispositions, as well as create an incentive for ethical conduct in development and innovation processes. The ethical fulcrum evolves around Løgstrup's Ethical Demand and his notion of spontaneous life manifestations. From this, three ethical stances are developed; apathy, sympathy and empathy. By exploring both apathetic and sympathetic views, the ethical reflections are more nuanced as a result of actually seeing the user experience simulated through different user dispositions. Exploring the three ethical stances by visualising real use cases with the technologies simulated as already being implemented makes the life manifestations of the users in context visible. We present and discuss how animation-based sketching can support the elaboration and examination of different ethical stances towards the user in the product and service development process. Finally we present a framework for creating narrative representations of emerging technology use cases, which invite to reflection upon the ethics of the user experience.

Categories and Subject Descriptors
H.5.2. [**User Interfaces**]: Evaluation/methodology, Prototyping, User-centered Design.

K.4.1 [**Public Policy Issues**]: Ethics

General Terms
Design, Experimentation, Human Factors

Keywords
Animation, sketching, user experience design, ethics, RRI, scenarios, design thinking, løgstrup

1. INTRODUCTION
This paper discusses how animation can be applied to simulate future applications of the designs to elaborate and examine different ethical stances towards the users in the product- and service development process through an empirical study of Responsible Research and Innovation (RRI) in industry cases. The challenge of every design and innovation process is to designate as well as reflect upon what this particular innovation will bring into the world; how it will change practices, perceptions, and relationships [1]. The common dissection between invention and innovation is that the latter not only creates something new, but in fact changes the way people live [2]. And with this change comes responsibility and ethical challenges for the designer. In the wake of these challenges the need for responsible research and innovation enters the picture.

The authors recognise RRI as a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products in order to allow a proper embedding of scientific and technological advances in our society [3]. RRI has mostly been used to determine methods and frameworks for inclusion in publicly funded research programmes across Europe. However, in industry, there are few active incentives for companies to innovate responsibly, and even fewer methods by which such incentives might be implemented. This is emphasised by the recent call for new knowledge to create this link between research into responsible innovation, and methods for the industry [4] [5].

This need for industry incentive contrasts the movement within the field of design thinking. Throughout the last decade, design and designerly ways of thinking about and acting upon the world, has gained widespread popularity [6] [7]. The movement towards a user-centred design approach, pioneered in the late 1980's and 1990's [8] [9] [10] has given rise to later years emphasis on the concept of 'user experience' [11] as the common denominator for the end-goal of all user-centred design processes. User experience design as an industry-oriented praxis details the need for understanding and testing the user's experiential quality, when developing new products and services [11].

Until recently however, the user experience design discourse lacked a discussion of the ethical dimension underlying its approach. At an earlier ETHICOMP conference, Vistisen & Jensen [12] presented a framework discussing the notion of user experience design from an ethical point of view. Showing how the notion of 'user experience design' creates an underlying responsibility for the designer. Designers claiming to be user-centred or to be designing in the context of the user experience also implicitly commit to shape and form certain aspects of the experience of a group of human beings (ibid) - thereby adopting

part of the responsibility for these experiences and their consequences. This experience can both be a small flutter in the user's way of performing a simple task enhanced by a given design, or it can be a life changing experience, brought about by an all-encompassing design strategy which catapults the user out of his everyday life [11].

With Løgstrup's [13] 'ethical demand' as their fulcrum [12], the user becomes 'the other person' to the designer in the design process. Categorising different ethical stances towards the user, they create a framework consisting of: apathy - the strict adherence to rationalism, sympathy - the reaction to an effect, and finally, empathy - the reaction to a cause. While the framework gave rise to intriguing discussions, no aim was given at the time as to which empirical domains this framework could be applied, neither which methods might enable the user-centred designer to actually explore the different user stances in real world settings.

Pairing RRI with Vistisen & Jensen's ethical perspective on user experience design creates a fitting industry oriented framing of how actors might form co-responsible relationships. While RRI first and foremost asks what kind of future we want innovation to bring into the world, ethical user experience design challenges us to discuss the underlying user dispositions during these innovation processes. In this paper we narrow the discussion down to focus on a certain design approach, animation-based sketching, by raising the question: how can animation-based sketching support the examination of ethical stances towards the user in the product and service development process? The next section will elaborate the ethical framework used in the exploration of the different user stances.

2. THE ETHICAL DEMAND

Løgstrup's ethical demand [13] differs greatly from other normative ethics [14] through its ontological and situational approach to ethics. Thus, a framework for a design process based on the ethical demand will always have to be user centred and situated. The core concept of Løgstrup's ethics depend on the dyadic meeting, where the 'I' (the designer) is responsible for acknowledging the unspoken ethical demand posed by 'the other person'. In the design process, and whenever a design is used, the 'I' will be the designer, while 'the other person' will be the end-user of the design. The design itself is mediating the dyadic meeting. The unspoken demand itself consists of the so-called life manifestations like mercy, trust, a plea for non-violence, and the openness of speech, among others [15] [16]. In Løgstrup's thinking, the 'I' has a responsibility to bring out the full potential of 'the other person's' being by acknowledging and respecting the unspoken ethical demand in their meeting [17].

Not only does the demand pose a considerable responsibility on the designer, but on the design process and the design itself. The dilemma of the unspoken ethical demand becomes apparent when it is turned *"(...) into an outward, manageable principle that is supposed to be able to operate as a magical principle and solve all problems. The result is that the demand becomes nothing but a cliché."* [13] It turns into a cliché, because the 'I', as the designer, will be the one who solely articulates and sets the conditions for the meeting, taking his knowledge and the existing rules and systems into account, without acknowledging the life world of the user, 'the other person'. This is, what Vistisen and Jensen [12] call the apathetic ethical stance toward the user. The user is just a means of input for the intended end, the final design.

To avoid this, Løgstrup emphasises the need for doubt and uncertainty on the side of the 'I', the designer in our case, since "(t)hinking and imagination become equally superfluous. Everything can be carried out quite mechanically; all that is needed is a purely technical calculation. There is no trace of the thinking and imagination which are triggered only by uncertainty and doubt." [13] Only by constantly questioning oneself, the designer can ensure a certain, needed openness toward the design process as well as the users involved.

Still, the designer needs to acknowledge his ethical responsibility as a designer. Meaning, it is important for the designer to make necessary choices in the design, to not only sympathise with the user, giving him whatever he demands. Instead, an empathetic design approach needs a deep understanding of the life world, which comprises not only of the tacit, but also of the systemic knowledge. In this, the three ethical approaches to design should be regarded as steps in the design process, especially when paired with a flexible and changeable method like animation-based sketching. As our case analysis will show, the design team uses all three ethical stances to accomplish a design concept, which takes the life world of the end user as well as the given task and the systemic needs into account.

3. RESEARCH METHOD

Experimenting with the pairing of RRI and ethical user experience design, we facilitated a three-week-long innovation workshop called U-CrAc, an abbreviation of User-driven Creative Academy. This workshop format originates from the LUDINNO research project, which was founded by The Nordic Research Council [18]. The objective of LUDINNO was to establish collaboration among participating companies and consultants with students and researcher through playful user-oriented laboratories or learning labs. From the perspective of the university this was an initiative to engage in the role of the civic university as there within the associated academics was a fundamental interest in knowledge application within the surrounding society. However the intention was not to take a subservient role, but instead engage as an influential actor and equal partner in a Triple Helix constellation with industry and government. The Triple Helix constellation builds on the idea of synergy between involved partners as; *"Industry operates in the Triple Helix as the locus of production; government as the source of contractual relations that guarantee stable interactions and exchange; the university as a source of new knowledge and technology, the generative principle of knowledge-based economies"* [19].

U-CrAc, has undergone several changes, as the workshop design itself is an iterative process in which we, the educators and researchers, seek to explore new methods and techniques. U-CrAc builds on the pedagogy of Problem-based Learning, and each of the 22 groups was given an assignment with an elaborated problem. These assignments had a combination of IT, experience and health dimensions and was provided by both local companies and public organisations, which in the following will be entitled clients. Throughout the workshop there is an on-going collaboration between the students and the associated client.

The workshop is divided into three phases; Fieldwork, Ideation and Concept development. Each phase had a dedicated week: the students performed ethnographic user studies in the first week and interpreted the observations into what we phrase *innovation tracks.* These innovation tracks became the starting point for the following idea and concept development process, which is the empirical focus of this paper. In these phases, Ideation and Concept development, the design students goal was to both explore new ideas as well as anticipate how these new ideas might affect the user experience. Furthermore the students were tasked

with exploring their ideas in different animation-based sketching formats, which opened up for different types of ethical reflections.

The students were instructed to use various forms of animation-based tools to help the companies simulate and reflect upon how different ethical stances towards their users could potentially affect the user experience of their product or service proposition. Through the workshop we examined how these methods could be used as a foundation for the participating companies to explore and experiment with the desirability and feasibility of their upcoming pipeline. Establishing an ethical point of reflection early in the process might affect the users final experience. Later in this paper we will take a deep dive into one of these innovation cases, deconstructing how animation-based sketching was used to explore multiple user-dispositions, and assess their ethical stances in regard to apathy, sympathy, and empathy.

3.1 Using animation as tool to sketch ideas

A method was required for the design students to express and externalise the different ethical stances towards the users in their ideas. Previously film scenarios has been used to externalise experiences through time, and in context as pointed out by Raijmakers [20] *"film is definitely the most powerful tool to an emotional understanding of the user"*. Furthermore, the linearity of video creates a constrained narrative, which may become an agent for change, functioning *"...as persuasion to present complex ideas in a concentrated and exciting way for influencing research directions and decisions,"* [21].

Despite its previous uses in design, and innovation processes, video as a sketching medium is by default limited to capturing the world of what is, and is only able to illustrate the world as it might be when the scenario is representational through existing artefacts. But, when concerned with expressing challenges regarding emerging technologies, and anticipate and reflect upon the possible user dispositions around these technologies, video simply lacks expressiveness. Our hypothesis was that exploring possible user dispositions in new and innovative contexts required a design material in which the designers would have a larger degree of control of the simulated use case for an idea. Such a potential was found in *"...the full transitional control of the subject matter"* in animation [22]. Animation can be defined as "the process of generating a series of frames containing an object or objects so that each frame appears as an alteration of the previous frame in order to show motion" [23]. Further, animation represents an abstraction of reality [24], and as a temporal 4-dimensional medium [25], it is able to simulate qualities such as movement, flows, transitions and timing from not-yet existing artefacts [26].

The use of animation as a tool to explore new design possibilities has previously been explored by creating animated use cases to gather feedback, and to explore the fuzzy front end of design ideas [27]. Similar studies were accounted for in Fallman et al [28], Fallman & Moussette [26] and Bonanni & Ishii [29] who used stop motion animation in early digital and architectural design processes. Others have used animation to augment traditional film [30] [31] [32] [33]. Despite being widely used, this approach in general does not address which qualities of animation actually makes it suitable in the design process. The techniques themselves are not examined in detail. Vistisen & Poulsen [34] investigate this dilemma in greater detail and assess that the simulative nature of animation enables the designers to create strong narratives, in which new technologies can be integrated into a believable use-context. The use of animation in this paper echoes this approach, by not emphasising the specifics of the animation techniques themselves, but rather by experimenting with animation as the enabling technology of exploring user dispositions in RRI cases. However the goal is not to create specialised tools either, as is the case with recent contributions [35] [36] [37]. Instead we place animation as a broad set of techniques, with a broad set of existing tools, that may be feasible to apply in the exploration of designs that does now yet exist - or in other words, to address the 'what if...' questions of RRI [38].

3.2 Selection of workshop case for analysis

To record the design students animation-based sketches we used a participant-generated web-platform [39] as a modified type of a technology probe [40]. The web-platform provided a common frame of reference for the facilitating researchers, the participating companies, and the design students to discuss, and reflect upon the different stages of the ideas, and ultimately the different user-dispositions inherent in each of the ideas.

From examining the sketches a general insight was how the multitude of animation-based sketching methods all seemed to enable the creation of sketches, which explored ethical user stances from the Løgstrup-based framework. Furthermore the explorations in general adhered to the primary concerns of RRI described by Stilgoe et al [41] as anticipating technological emergence, reflecting upon it's consequences, inclusion of stakeholders, and responsiveness towards the next step. However, dependent on the industry case, it was also evident that some of the produced sketches explored a broader range of ethical user stances than others. While the RRI perspectives can be identified as a higher meta-level aim to shape, develop and align existing and future technological innovation in the process [42], the three ethical user stances from Vistisen & Jensen are more evident in the details of the sketches. Thus, to further assess how animation-based sketching enables us to explore user dispositions in RRI cases, we selected one of the cases which explored aspects of all three ethical user stances for a further case study.

The selected case was a collaboration between the retirement home 'Plejecenter Lykkevang' and the Danish health care innovation center 'Copenhagen Living Lab'. The case challenged the students to explore how to engage and empower elderly residents in smart retirement homes. The students' ethnographic field studies were captured as a series of four video segments showcasing the limited focus on creating activities for the still-active residents at todays retirement homes. The video material produced helped the design students to map the current apathetic situation, and provided a basis for the students initial statement of the 'right design' [43]: *how can we support the activities of the elderly by creating scalable social experiences which motivate both physical and social activity?*

From the mapping of the current state of the retirement homes the design students began their ideation process, and sketched their ideas into scenarios [44]. Through video enactments and by applying animation techniques and effects these scenarios became visualised as a series of animation-based video sketches. The next section presents the produced sketches, and reflects upon the user dispositions the sketches portrayed.

4. CASE ANALYSIS

A total of three initial animation-based video sketches were made before the design students arrived at the final concept of the 'PlejePad'.

4.1 The interactive experience room

The first concept generated was the interactive experience room with projected visualisations on the walls, aimed at creating an

immersive environment for the elderly to experience without having to travel to other locations other than a designated area of the retirement home [45]. In the sketch, we see the caretakers help the residents into the experience room followed by a series of different content types, the elderly would be able to experience inside the room (Figure 1). The sketch uses green screen video recordings with animated motion graphics overlays to simulate the digital walls of the experience room.

While the simulated interactive environment would seem to solve parts of the design problem of creating a social experience it is evident in the use case how the concept actually shows an apathetic user stance. The elderly are placed inside the experience room by the caretakers, and are then left for themselves to experience the content. While this may create an experience in by terms from [11], the experience really does not in any way solve the underlying problem of the elderly needing more social and active interactions in their daily routines. Instead, the elderly are treated as a component in a procedure of being placed inside an installation, receiving a designated dose of stimulus, and are then left to their normal routines again. Thus, the scenario helps to clarify how the use of digital design does not necessarily result in a solution which actually solves the problem, but might as well become an extension of the existing apathetic situation in the system of the retirement home.

4.2 Digital games in the common area tables

The second concept seeks to create a social and active experience for the elderly through digital games integrated in the common room tables [46]. The simulated use case illustrates how two residents activate the table after dining together, before choosing between a range of classic board games in a digital format (figure 1). The scenario is made by animating a series of timed keyframe animations on top of the table to simulate the digital interface and games.

Figure 1. The interactive experience room (top) and the digital table games (bottom).

In the scenario we see how the elderly are able to interact via gestures in order to navigate the digital interface of the game table. Furthermore we see how the table mediated the social interaction between the two participants. However, the scenario also showcases a user disposition in which 'the need for a social and active experience' is literally translated into playing a game together. While the idea of an interactive dining table is novel, the scenario does not show how the technology helps the elderly become better suited to engage in active and social activities. Neither does the solution empower the elderly to take control of the experiences, besides giving them the opportunity to sit and play predesignated games. In this regard, the scenario explores a 'sympathetic' user stance by showing how a seemingly novel solution to the problem actually only treats the symptoms and not the cause for the problems with lacking social and active daily routines at the retirement home. Thus, the technology is paired with the person, but not recognising the contextual setting or underlying motivations for the problem faced in the context.

4.3 Social touch screen in the living room

Following the first two sketches exploring possible apathetic and sympathetic user experiences at the retirement home, the students were able to reframe the problem into: *how the activities of the elderly can be supported by integrating social and active experiences into their existing daily routines?*

Through this reframing, the third animation-based sketch explored the use of a social touch screen system in the individual apartments of the retirement home [47]. The sketch shows a scenario with a resident establishing a video chat with another resident, arranging a social activity in the common rooms (figure 2). The interaction with the touch screen is simulated through simple stop motion animations.

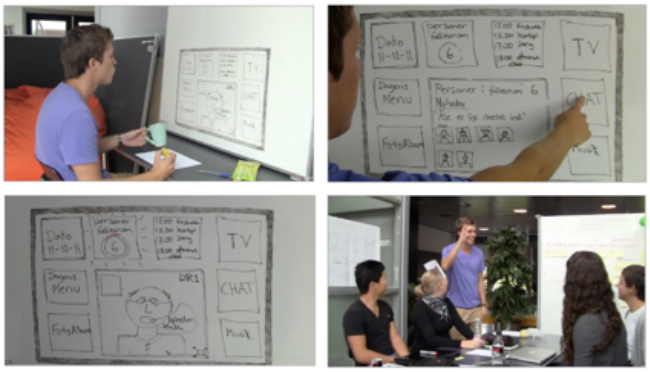


Figure 2. The early vision behind PlejePad, depicted as a social touch screen system in the living room.

Through this scenario the design students explored how to establish a more empathetic user stance towards enabling the elderly to actively view and manage the social activities through a device located in the context of the apartment. The empathetic disposition is evident in the idea's focus on taking the current living situations of the elderly as the starting point of concept, further elaborating how the new device can tap into the daily routines, and make it easier to communicate and participate in activities at the retirement home.

Through making the sketch the design students realised that even though the general aspects of the idea addresses the cause for the problem of inactivity and lack of social interactions, the touch screen solution might not fit the digital literacy of the majority of the elderly residents, they had met during their field work. The touch screen was a product of the design students current understanding of the technological landscape, and did not accommodate the same level of empathy as the overall idea about using a screen in the apartment to mediate the social activities for the elderly. This reflection upon the ethical stance towards the literacy and social fit of the concept, led to the reframing towards the final idea of 'Plejepad' (english: NursingPad).

4.4 The PlejePad concept

The final animation-based video sketch makes use of a range of animation techniques to simulate the screen-based 'PlejePad' [48]. The concept is a smart TV system, which is controlled through a

traditional remote control, adhering to the technological literacy of a medium and interaction device most of the elderly are familiar with. Furthermore, the use of animation is used to integrate the prior insights about the apathetic user disposition of the situation as it is at the retirement home. By animating a clock in the top left corner, and running a fast-forward time lapse of the daily routines of the elderly persona, it is illustrated how the elderly often is confined to be sitting alone in the apartment, often in front of the TV (figure 3).

Figure 3. The apathetic situation of the current daily routines of the elderly, depicted via animated annotations.

The apathetic user disposition is illustrated in a quick and straight-forward manner by using easy to understand visuals to emphasise the narrative setting and context of the problem. This helps to establish a clear connection between the apathetic status quo, and the following sequence in which the empathetic user stance is explored through the new concepts, integrating the exact same context and routines, but altered by the system's social mediation. The sketch makes use of keyframed interface animation to showcase how the elderly persona interacts with the system (figure 4).

The sketch shows how the proposed concepts acknowledges the cause of the problem, and circumvents it by making the TV the main hub for arranging and controlling social activities. The concepts thus takes an empathetic user stance in showcasing how a new emerging technology (smart TV systems) may be appropriated into a specific context (apartment in a retirement home) fitting the routines and literacies of the user. To explore the potential user experience of this empathetic stance towards the elderly persona in the sketch, the design students set up a concrete user scenario through a narrative of the persona 'Ole' interacting with his friend 'Helge' through the PlejePad system, arranging to participate at a social activity at the retirement home (figure 4).

The scenario illustrates how Ole communicates with Helge through the voice and voice-to-text messaging service ind the system, coordinating to participate in an activity shown in the 'Daily overview' function in the system. After agreeing upon the activity, Ole goes back to his daily routines in the apartment, until the TV system gives him a reminder about his appointment with Helge. When pointing the remote at the reminder, Ole sees which residents are present in the common areas for the activity, and makes ready to leave the apartment to meet up with Helge. The empathetic user stance is again evident in how the design students explored the integration of technologies such as peer-to-peer communication, online scheduling, indoor wayfinding, and intelligent assistants. The technologies integration into the context presents a way to solve the cause for the in-activity problem, while staying true to the literacies and routines of the person, and further empowers him to reach out and connect - augmenting the social sphere of the entire retirement home.

Throughout the final part of the animation-based video sketch, the design students explore how the system might adhere to the anticipatory function deemed important by the RRI discourse [41]. We see how the caretakers can customise and edit which apps and functions are available to the individual smart TV, which shows how responsibility can be delegated between the industry stakeholder (retirement home) and the end-user (the elderly).

Figure 4. The Plejepad system in the living room (top), the personas interacting with the system (middle), and the back end customisation features (bottom).

The animation-based video sketch takes the viewer through a narrative in which we get to explore the apathetic status-quo of the present situation, and is guided through a story of the elderly persona, as the sketch builds up its case for how an empathetic user experience can be achieved. Through its narrative structure the division of touch points between the elderly, the caretakers, and the context of the retirement home are explored, and the inherent responsibilities are made visible. In tandem the sketches, exploring apathetic, sympathetic, and empathetic user stances towards integrating digital technologies into the problem domain, invites the viewer to reflect upon both the application of a certain technology, as wells as the implications it may have for the user experience of the involved stakeholders. The narrative format, and the use of animation to simulate the emerging technology, and modify the context helps to include a broad range of stakeholders in the reflective process of evaluating both the technical concept as well as the underlying ethical user dispositions. Thus, animation-based video sketches becomes more of a reflective tool, than a communicative tool, as would normally be the case for animated narratives [22]. In the next section we will gather the insights from the case analysis, and present a possible framework for exploring user experiences with interactive technologies in animation-based video narratives.

5. A NARRATIVE FRAMEWORK

As we have seen in the case above, the exploration of ethical user stances is not necessarily a process of choosing one user stance, but more a flexible process of reaching an empathetic user experience as the end goal. By exploring both apathetic and sympathetic views, the ethical reflections of the stakeholders become more nuanced due to the process of actually seeing the user experience simulated through different user dispositions.

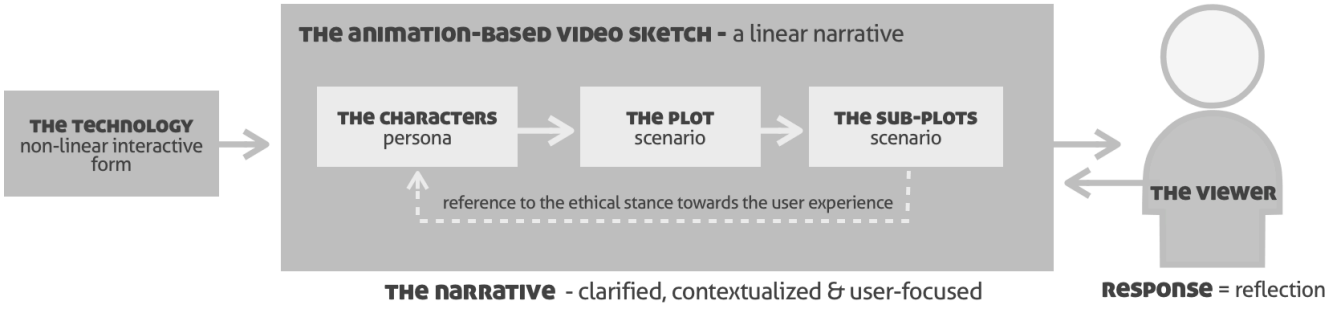


Figure 5: Framework for creating linear animation-based video sketches which explores new technologies from an ethical user perspective.

Exploring the three ethical stances by visualising real use cases with the technologies simulated as already being implemented makes the life manifestations of the users in context visible and relevant. Thus, through animation the scenarios are able to simulate how the ethical demand is applied in a given setup of users, industry stakeholders, and newly developed technologies. This offers a strong incentive for reflecting upon critical issues of how to create responsible innovations, since the user dispositions are explored in an easily comprehended format. Furthermore using animation shows to be a flexible set of techniques, which enables a broad range of cases to be simulated and communicated.

Systemising the functional components in the animation-based video sketches we get a framework through which an interactive technology is placed inside a linear scenario. Hereby a user persona acts as characters in a story, which takes place in a given context. The plot of this story revolves around any issues to which the interactive technology is presented as a possible solution. From the scenarios exploration a reference is drawn back towards the persona, illustrating which ethical stance the technology takes upon the persona in the given context and use case. The framework may be illustrated as figure 5.

Through the process of applying animation in a narrative format, which is not aimed solely at storytelling, but rather at creating ethical reflections, we get a framework for the construction of such animation-based video sketches. Using persona stand-ins for the real observed users [49] and placing them in a real world scenario [44] and by establishing a clear point of reference to how the technology affects the life world of the persona. Thus, the user disposition is made visible and inclusive for others to reflect and comment upon.

Concerning the practical feasibility of using animation to explore ethical user dispositions one might ask whether the techniques and framework are generically applicable. Considering the RRI discourse's emphasis on anticipation, reflectivity, and inclusion we argue that this question depends on the technological issue at hand. If we deal with more or less normative issues, like designing with existing technologies, and with existing design patterns [50] we might be less inclined to simulate the user experience in an animation-based video sketch. On the other end of the spectrum, fields like design fiction [51] [52] and critical design [53] recently have been proponents for speculating in future scenarios for both problems and contexts that are still unknown. Here, simulating and speculative prototyping is the only possible tool available. This critical domain of design has no normative qualities, but is quite often concerned with the speculative futurism, rather than the present world 'as it is'. Inside this spectrum, between the purely normative, and the purely speculative, we might place animation-based video sketching of ethical user dispositions as 'the middle ground'. Maintaining a critical perspective on new technologies and their applications, but with a clearly strategic aim to explore how the relationship between users, industry an R&D should be established to reach the 'right impact' [41].

Once you work in a narrative setting, focus is taken away from the design itself. Instead, context and world building, the conflict, and characters become important and present. A narrative is open for interpretation, enabling a discussion which surpasses mere functionality and the design as such. A narrative opens for possibilities, and engages the reader, viewer, listener. And with engagement comes participation and empathy. A deeper understanding of the design and its purpose and possibilities within the world. This exploration is not based on some far-future utopia or dystopia, but on how we make the most responsible user experiences in the near-future. Being able to simulate, and clearly articulate multiple user dispositions in such near-future scenarios is the main contribution of animation-based video sketching for RRI.

6. CONCLUSION

Through the research question of this paper we explored how animation-based sketching can support the elaboration and examination of different ethical stances towards the user in the product and service development process. By using the ontological ethics of Løgstrup as a framework for the design process we tested how the life world of the end-users could be taken into account, as well as how the designer could explore multiple user dispositions towards establishing an empathetic user experience.

As argued, working with the uncertainty prescribed by Løgstrup demands flexibility from the designer and the design methods put to use. Animation-based video sketching is a set of tools, which enable the designer to create simulated narratives of the near-future, to promote reflection upon the desirability and relevance of the user experience depicted. By exploring both the apathetic, sympathetic, and empathetic sides of the design problem, a more nuanced reflection can be achieved. By creating more operative deliverables for ethical reflection, the examination of the responsibilities between an innovation project's stakeholders may also become more inclusive.

We have presented animation-based sketching as a viable tool to create such operative images for ethical reflection upon the user dispositions when designing new interactive products. We

contribute to the existing discourse by showcasing how animation can be used to simulate the near-future use of new emerging technologies, and make their ethical user stances visible to both the viewer and the designer. Thus, the set of techniques, and the framework for their application in narratives as our contribution to the developing RRI toolkit.

7. REFERENCES

[1] Buchanan, R. 1992. Wicked Problems in Design Thinking. Design Issues Vol. 8, No. 2 (Spring, 1992) , pp. 5-21. MIT Press

[2] Chayutsahakij, P. and Poggenpohl S. 2002. *User-Centered Innovation*. Proceedings of The European Academy of Management 2nd Annual Conference on Innovative Research in Management EURAM, Stockholm, Sweden

[3] von Schomberg. 2011. *The quest for the "right" impacts of science and technology. An outlook towards a framework for responsible research and innovation.* in: (eds M.Dusseldorp, R. Beecroft) *Technikfolgen abschätzen lehren. Bildungspotenziale transdisziplinärer Methoden*. Springer

[4] ETHICOMP 2015 call for papers on RRI: http://www.dmu.ac.uk/research/research-faculties-and-institutes/technology/centre-for-computing-and-social-responsibility/tracks.aspx Retrieved June 30th 2015

[5] RRI Tools http://www.rri-tools.eu/project-description Retrieved June 30th 2015

[6] Brown, T. 2009. *Change By Design*. Harper Business

[7] Nelson, H. G., & Stolterman, E. 2012. *The design way intentional change in an unpredictable world*. Cambridge, Massachusetts: MIT Press.

[8] Suchman L. 1987. *Plans and situated actions : The Problem of Human- Machine Communication*, Cambridge University Press

[9] Greenbaum, J., & Kyng, M. (Eds.) 1991. *Design at work: Cooperative design of computer systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[10] Bannon, L. & Bødker, S. 1989. *Beyond the Interface: Encountering Artifacts in Use,* Aarhus University Press

[11] Hassenzahl M., Tractinsky N. 2006. *User experience - a research agenda.* Behaviour & Information Technology Vol. 25, Issue 2 2006

[12] Vistisen, P., & Jensen, T. 2013. *The Ethics of User Experience Design: Discussed by the Terms of Apathy, Sympathy, and Empathy*. I A. Gerdes, T. W. Bynum, W. Fleishman, S. Rogerson, & G. Møldrup Nielsen (red.), *ETHICOMP 2013 Conference Proceedings: The possibilities of ethical ICT.* Odense: Syddansk Universitetsforla

[13] Løgstrup, K.E. 1997. *The Ethical Demand*. Notre Dame: University of Notre Dame Press.

[14] Nyeng, F. 2000. *Etiske teorier*. København: Gyldendal Uddannelse.

[15] Løgstrup, K. E. 2007. *Beyond the Ethical Demand*. Notre Dame: University of Notre Dame Press.

[16] Løgstrup, K. E. 2014. *Etiske begreber og problemer.* Aarhus: Forlaget Klim.

[17] Pahuus, M. 1995. *Holdning og spontaneitet. Pædagogik, Menneskesyn og værdier.* Århus: Kvan.

[18] Edman, T. Keitsch, M., Vavik, T., Morelli, N., Poulsen, S. B., Koskinen, I., Holmlid, S., et al. 2010. *LUDINNO - Learning Labs for User-Driven Innovation.* Oslo, Norway: Nordic Innovation Centre

[19] Etzkowitz, H. 2003. *Innovation in innovation: The triple helix of university-industry-government relations.* Social science information, 42(3), 293-337.

[20] Raijmakers, R., Sommerwerk, A., Leihener, J., Tulusan, I. 2009. *How sticky research drives service design*. Service Design Network conference in Madeira. Chow, M. D. 1989. *The role of the video professional in a research environment*. ACM SIGCHI Bulletin

[21] Chow, M. D. 1989. *The role of the video professional in a research environment*. ACM SIGCHI Bulletin

[22] Stephenson, R. 1973. *The Animated Film*. Tantivy Press

[23] Baek, Y.K. and Layne, B.H. *Color, graphics, and animation in a computer-assisted learning tutorial lesson*. Journal of Computer-Based Instruction 15, 4 (1988), 131–135.

[24] Elish, M.C. 2011. *Responsible storytelling: communicating research in video demos*. *Proc. TEI'11*, 25–28.

[25] Vistisen, P. 2014. Abductive Sensemaking Through Sketching. Academic Quarter, vol 9.

[26] Fallman, D. & Moussette, C. 2011. *Sketching with Stop Motion Animation.* ACM interactions, Volume XVIII.2, March + April, New York, NY: ACM Press, 57—61

[27] Löwgren, J. 2004. *Animated use sketches as design representations*. interactions vol 11, issue 6, ACM

[28] Fallman, D., Zarin, R. & Lindbergh, K. 2012. *Stop Motion Animation as a Tool for Sketching in Architecture.* Proceedings of DRS 2012.

[29] Bonanni, L. and Ishii, H. 2009. *Stop-motion prototyping for tangible interfaces.* Proc. of the Third International Conference on Tangible and Embedded Interaction. ACM

[30] Mackay, W. 1988. *Video Prototyping - a technique for developing hypermedia systems*. CHI'88 Demonstration, ACM/SIGCHI.

[31] Vertelney, L. 1989. *Using video to prototype user interfaces.* SIGCHI Bulletin 21(2):57-61.

[32] Bardam, Bossen, Lykke-Olesen, Halskov & Nielsen. 2002. *Virtual Video Prototyping of Pervasive Healthcare Systems.* DIS 2002. ACM.

[33] Tikkanen, T., Cabrera, AB. 2008. *Using Video to Support Co-Design of Information and Communication Technologies.* Observatorio Journal, Vol 5. Obercom.

[34] Vistisen, P., & Poulsen, S. B. 2015. *Investigating User Experiences Through Animation-based Sketching.* The Motion Design Education Summit 2015 (MODE 15), Dublin

[35] Fernández J., Martens, J.B.O.S. 2013. *idAnimate: A General Purpose Animation Sketching Tool for Multi-touch Devices.* Proceedings of CONTENT 2013. IARIA.

[36] Davis R.C., Colwell B., Landay J.A. 2008. *K-sketch: a 'kinetic' sketch pad for novice animators.* Proc. Of Human factors in Computing Systems (CHI '08). ACM.

[37] Sohn,E.,Choy,Y.C. 2010. *Sketch-n-Stretch:sketching animations using cutouts*. IEEE Computer Graphics and Applications, vol. 99.

[38] Guston, D. 2013. *"Daddy, can I Have a Puddle Gator?" Creativity, Anticipation and Responsible Innovation*, in Owen, R., Bessant, J. & Heintz, M. (ed.) *Responsible Innovation: Managing the responsible emergence of science and innovation in society.* Wiley Press.

[39] Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin- Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., Eiderbäck, B., Lindquist, S., & Sundblad, Y. 2003. *Technology Probes: Inspiring Design for and with Families*. In CHI 2003 (17-24). ACM.

[40] Web 1: www.ucrac.dk. Retrieved June 30th 2015

[41] Stilgoe, J., Owen, R. Macnaghten, P. 2013. *Developing a framework for responsible innovation*. Res. Policy 42: 1568–1580

[42] Stahl, Bernd Carsten. 2013. *Responsible research and innovation: The role of privacy in an emerging framework*. Science and Public Policy 40 (6)

[43] Buxton B. 2007. *Sketching User Experiences - getting right design, and getting the design right"*, Morgan Kaufman

[44] Carroll, J. 1995. *Scenario-Based Design: Envisioning Work and Technology in System Development*. John Wiley & Sons

[45] Web 2: Interactive experience room. https://www.youtube.com/watch?v=26F6qOf4JfY. Retrieved June 30th 2015

[46] Web 3: Digital games in the table. https://www.youtube.com/watch?v=Q11VjTIbYPU. Retrieved June 30th 2015

[47] Web 4: Social touch screen. https://www.youtube.com/watch?v=tQ1JXiofFqU. Retrieved June 30th 2015

[48] Web 5: Plejepad. https://www.youtube.com/watch?v=K3Iy12y-kYs. Retrieved June 30th 2015

[49] Cooper, A. 1999. *The Inmates are Running the Asylum: Why High-Tech Product Drive Us Crazy and How to Restore the Sanity.* Indianapolis: Sams

[50] Gamma, E., Helm, R., Vlissides, J. 1994. Design Patterns. Addison-Wesley Professional

[51] Bleecker, J. 2009. *Design Fiction: A short essay on design, science, fact and fiction*. Future Lab., vol. 29

[52] Knutz, E., Markussen, T., and Christensen, P. 2013. *The Role of Fiction in Experiments within Design, Art & Architecture.* presented at the Nordic Design Research (NORDES 13)

[53] Dunne, A. 1999. *Hertzian tales: electronic products, aesthetic experience, and critical design*. London: RCA CRD research publications

Distorted Usability Design in IT Tendering

Kimmo Tarkkanen
University of Turku
Rehtorinpellonkatu 3
20500 Turku Finland
kimmo.tarkkanen@utu.fi

Jani Koskinen
University of Turku
Rehtorinpellonkatu 3
20500 Turku Finland
jasiko@utu.fi

Ville Harkke
University of Turku
Rehtorinpellonkatu 3
20500 Turku Finland
ville.harkke@utu.fi

ABSTRACT
Request-for-proposals (RFP) are documents in IT tendering that define the selection criteria, evaluation procedures and system requirements including system usability. IT vendors' perspective on RFP-originated system configuration and usability design is less studied than of IT procuring organizations. Analysis of empirical data collected from large IT tendering shows that from the vendor's perspective the objectives and means of usability design during the tendering differ drastically from general usability work. During the tendering, the fundamentals of usability recommendations can be based solely on the requirements of RFPs with no adequate intention to improve system usability in the use context. An ethical analysis of the situation and possible futures and alternatives is represented.

Categories and Subject Descriptors
K.4.1 [**Public Policy Issues**]: Ethics

General Terms
Human Factors

Keywords
Tendering; Request for proposal; Usability; Ethics

1. INTRODUCTION
Public information system purchases are subject to tender, for example in European Union member states and many other countries, if a certain monetary limit is exceeded. Request-for-proposals (RFP) are documents that begin a public tendering process – yet a common practice also in the private sector to run bidding between IT vendors. RFP defines a set of desired system requirements and a selection criterion for the proposed systems. As usability can dictate the success or failure of information systems, RFPs include also requirements related to usability. For example, RFP may define how usable the system must be, what the specific usability levels at minimum are, how usability must be designed by the IT vendor and how it is evaluated by the procuring organization as well as how much weight usability is given in the selection process.

In the light of recent research, usability requirements in the RFP should be addressed as performance measures [1]. A performance measure is e.g. how easily and quickly the user should be able to accomplish a certain task. Performance can be measured in terms of effectiveness, efficiency and satisfaction i.e. the elements of the definition of usability [2]. Other types of measures, such as the number of required usability design iterations, are not valid, reliable and adequately comprehensive to ensure the usability of the selected system [1].

In the mid-90s, very few RFPs specified HCI issues or usability activities beyond general descriptions [3]. Since then, the majority of the HCI studies on RFPs have concentrated on the procurement organizations to instruct them about best-practices in RFP creation [1, 4, 5, 6, 7]. Still, for example, a recent HCI workshop seeks for the scarce examples of "applying a human-centered approach in an effective way in government system procurement" [8].

However, we recognize that the research from the IT vendor's perspective on usability issues in RFPs is lacking. In this paper, we grasp the vendor's perspective with an empirical study on large-scale IT procurement. First, we describe what kind of problems usability designers and evaluators working for the IT vendor and improving the system in competition confronted during a tendering process and how they handled and responded to usability requirements defined in the RFP. Next, the problems and possible consequences are analyzed from three most noted ethical points of view, namely (Kantian) deontology, utilitarian consequentialism and virtue ethics (see [9]). Finally, a practical advice for usability design and evaluation for both parties is presented.

2. EMPIRICAL DATA COLLECTION
Methodologically the study follows the tradition of action research [10] flavored with participatory/ethnographical data collection practices, where researchers intervene in actual (design) work, introducing their contribution to the work setting, and collecting information through observations, interviews and reflecting their own experiences while the intervention is applied and evaluated in the case organization. In this study, the empirical data was collected from large-scale IT procurement, in which two researchers were partaking in the usability design and evaluation practices of the IT vendor. Our task as usability experts and researchers was to (quickly) improve the usability of the proposed systems before the procuring organization began its own usability evaluation.

We pre-evaluated usability of two interconnected enterprise-wide systems (Alpha and Beta), one web-based front-end module (Gamma), which was to be integrated with Alpha and Beta, and one separate system module (Delta) of another work domain, which was also later to be connected with the above mentioned systems. The systems were off-the-shelf products, live in several

organizations, yet configurable to specific needs of the procuring organization. Thus, systems can be compared with any configurable ERP software at the market. All the tested systems required and were allowed to have some configuration to meet the specific requirements of the RFP. Naturally, the RFP prevented the total automation of test tasks, but normal localizations and modifications were allowed.

The evaluation procedures were described in detail in the RFP, which allowed us to copy and follow the evaluation practices of the procuring organization in advance. For example, both Alpha and Beta had well over 10 usability test tasks described and to be performed by the end-users of the purchasing organization. Also other usability evaluation methods were to be applied in the tendering although we concentrate here mainly on findings from usability testing and expert reviews. According to the RFP, usability measurements with various weights were to be quantified to numeric values and total points given to a system in competition. Usability points had high significance for the decision.

Usability of Alpha and Beta were tested with four potential end-users in one-by-one test sessions each lasting two hours. The number of usability problems found and analyzed by two test administrators was 29 in Alpha and 23 in Beta. In addition, Beta was expert reviewed with cognitive walkthrough by one usability expert resulting 11 usability problems reported to the IT vendor. The Gamma system was evaluated by three usability experts by following the usability test tasks described in the appendix of the RFP. Due to Gamma's target audience, experts could take the role of prospective users as well. Gamma was reported having 30 usability problems. Usability of the Delta system was not evaluated by prospective end-users. Evaluation of Delta was conducted in 2 hours' usage demonstration given to two usability experts by the system developers. Usability remarks and related refinements (approx. 10 pcs) were created on-the-fly during the demonstration.

All evaluation sessions were audio recorded and the sessions with prospective users were also video recorded. The recordings were, first of all, used for writing evaluation reports to the IT vendor. Therefore, recordings were not transcribed or encoded for deeper analysis. The data we present in this paper is actually produced by us as usability designers/evaluators during observing prospective users and systems and writing usability test reports of the sessions. IT vendor representatives have not been interviewed for this study. We are also unknowledgeable about how (and how many of) found usability problems in our studies were actually fixed by the system developers although we know that the feedback from developers was positive in general. Nevertheless, in this paper, we aim to characterize the nature of found usability problems and especially highlight some of the proposed solutions and techniques as a response to these problems. Not all the found problems and recommendations (73 in total) are represented, but few selected ones that we consider as highlighting possible ethical conflicts due to tendering situation. Thus, this re-analysis of the problems, solutions and results of our own usability work has not been encoded in any systematic way, because our aim is to "tell the story", like ethnographers do, about usability design during the tendering process and raise discussion of its ethical nature.

3. USABILITY DESIGN IN THE CASE

Usability evaluation procedures and criteria represented in the RFP could be described as rigid. However, the RFP did not address the exact measures of usability (e.g. how efficiency is measured) at the time the usability evaluation and improvements of the systems needed to take place. Although it was rather easy to construct such operationalization of measures with previous knowledge of usability evaluation, this minor lack of information was a trigger to please and think like the usability evaluators of the procuring organization. The main target shifted from improving the system usability for the sake of end-users and their work to improve the system performance against the published measures (which needed to be figured out first). In other words, instead of fitting the system to the needs of the end-users of the organization, the tendering situation enforced us as usability designers to streamline the absolute efficiency and effectiveness of the system while also considering what another usability expert would think or favor in certain human-computer interaction situations.

In the Alpha and Beta user tests, only minor proportion of the test tasks were fully accomplished by the participants in a given time frame. The task completion rate, a common measure of effectiveness, was one of the measures mentioned in the RFP. As a consequence of failing in the task execution, we note that other evaluation criteria were also poorly met: The number of errors and interaction steps increased (more mouse clicks) and user satisfaction decreased. Thus, usability designers' main goal from the vendor's perspective became first to ensure that users would complete tasks, and second, to complete those with the most minimum effort (e.g. minimize steps of interaction).

We found some test tasks as far more important to accomplish as others, because of implicit and explicit connections between the tasks. Quite common test setting is that if the user fails to perform the task 1 (e.g. login to the mail system), she cannot perform task 2 either (e.g. send email to John). On the other hand, we noted for example that if the user succeeded in the task 4, she would accomplish also the task 7, because she saw certain information or applied a certain feature already in the previous task. Thus, it was more important to put quick design effort on task 4 than on task 7. While learnability and memorability are indeed important aspects of system usability, here the importance of the tasks i.e. which task needs the most redesign effort was based on pure mathematics not on its importance at real work.

Use scenarios and (flows of) usability test tasks in the RFP constituted a rigid description of the end-users' work. These work flows (e.g. the sequence of usability test tasks 1 to 5) were not always found the most optimal for the systems' logic. For example, task A before B would have been an ideal sequencing for the system (Delta), however, the RFP requested the sequencing of task B before task A. Although both task sequences pursued for the same work goal, we had to figure out new efficient paths to achieve the best evaluation result. It must be remembered that the rigid RFP could not be overridden by "better solutions", because winning the tendering would depend on the system performance in pre-defined and pre-ordered test tasks. On the other hand, we claim that the RFP may have not included an in-depth understanding about the current and desired work practices of the end-users. Also communication limitations during the tendering seemed to effectively hinder the IT vendor fixing the deficiencies of the underlying assumptions of the "best practices" and introducing alternative task flows. Figure 1 visualizes the overall usability design situation during the tendering by describing the (non-)overlapping sets of the requirements and objectives of the tendering parties.

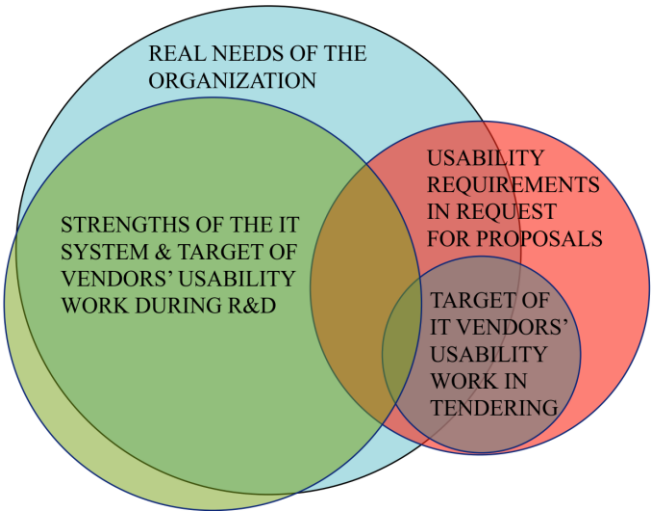


Figure 1 Depending on the RFP an IT vendor's usability work can be differently focused than in traditional R&D [11].

The task completion rate and efficiency of the tested systems were improved with several practical redesign recommendations addressed below (selected and not in a particular order). We recommended to 1) add extraneous shortcuts and links to guide users to the desired navigation path. For example, we note that "finding navigation path to X is too hard for the users, therefore add Y [menu item or link] that directly navigates to X". This seems plausible change within the normal system configuration that speeds up the execution of the tasks, although we had no knowledge of the appropriateness and usability of the resulting information architecture[1]. Neither, considering the tendering objective, we should care as we note that "the user does not have to know 'what happened' or 'where I am' [after following the shortcut], because users are presumably able to continue to the next test task". So, while the objective to improve system usability (efficiency and effectiveness of operating the system) is a common practice, the recommendations are somewhat contradicting with general usability guidelines. For example, the notion of users above clearly conflicts with the visibility of system status in Nielsen's heuristics [12], which guides the system design to keep users always informed of what is going on.

Further, we recommended to 2) modify the default settings of different views, frames and menus to provide more timely visibility with the flow of the test tasks. For example, testing the Beta, we note that "the number of interaction steps is minimized in the test task, if the window frame X is open by default when the application is opened". The redesign was possible, because the system frames were customizable, yet it remains unknown whether users favor to begin with the frame open or close. Next, in order to diminish interaction steps in data selection and data searching tasks, we recommended to 3) modify default data items and their order e.g. in drop-down menus and lists for specific data to appear more "automatically", and 4) to change terminology used in buttons and menu titles etc. to correspond with the terminology used in the test tasks. A common rule of thumb in planning usability test tasks is that these are not formulated in terms of the application (yet with terms users are familiar with), in order to avoid giving too much guidance (see [13]). In the opposite manner, the system redesign here could take advantage of the terminology of the test tasks. On the other hand, the terminology used in the system needed also to be changed due to the formulation of test tasks. For example, test tasks 5 and 6 were formulated with the word X, which encouraged users to try the X-function of the system. However, the X-function did not support the execution of the tasks and thus the function title needed to be changed. Again, we did not know whether the wording of the test task is appropriate in the real use context.

We also recommended to 5) replicate pieces of information in different views and frames to guarantee information visibility and access, 6) to inactivate features (e.g. links, buttons) in order to force users to follow a desired navigation path and to avoid error messages (which could appear due to unimplemented features) and 7) to avoid showing error messages and pages (e.g. by inactivation) because users may get frustrated and experience the system as erroneous. Replication, inactivation and avoiding error messages are not common means in the usability designers' tool-box. Quite the contrary, as Nielsen [12], for instance suggests following minimalist design, letting users be in control and helping them to recover from errors with informative dialogs and emergency exits.

In summary, the above examples of different solutions to improve task completion and task performance efficiency (1-7) are solely based on the requirements of the test tasks and not on real users' real needs, desires or the requirements of the actual work tasks. We don't know whether, for example, opening a frame by default is annoying real users or is beneficial at all in their daily work. Although such usability re-designs are possible for the system demonstration and testing purposes, such a system configuration may not be feasible during the real use of the system after the tendering. Even if technically possible, that may not reflect the best system-organization fit denoting poorer efficiency and effectiveness of the system than it was during the tendering. At the worst, the RFP followed redesigns are not implementable at all in the future configuration or turn out to be catastrophic in the real tool use situation.

What we know is that, for example the change in the frame default setting will (possibly) help users to recognize the required information and accomplish the test task. While such change may dissatisfy users also during the actual usability test performed by the procuring organization, users would certainly be at least equally dissatisfied if they failed in task accomplishment. Moreover, in this tendering case the weight of user satisfaction was only at mediocre level compared with the weight of other usability measures in the RFP. Thus, arguments to improve the absolute efficiency and effectiveness of the user interface against the RFP metrics irrespective from the use context were strong. However, that is streamlining the system that should not be streamlined yet. From the purchasing organization perspective, they would not evaluate systems that are 'as-is' (i.e. off-the-shelf products without modifications or local configurations) nor 'to-be' (vendor's best attempt to solve the problems with IT), but on the fly assembled shaky systems, that possibly jeopardize the fundamentals of the procurement policies. Next, the case findings and possible implications in IT tendering are further analyzed from the ethical point of view.

[1] Consider that shortcuts are always created for all the required tasks instead of organizing and grouping them e.g. based on usage frequency. Perhaps piles of shortcuts in your PC desktop look such a mess!

4. ETHICAL ANALYSIS

In the ethical analysis, we refer to three ethical points of view, namely (Kantian) deontology, utilitarian consequentialism and virtue ethics (see [9]). Deontological (duty) ethics is a normative ethical approach, which values an act to be ethical or not based on rules or duties, which should be obeyed. Probably the most famous and noted deontological ethical theory is the Kantian categorical imperative, which could be stated as 'an act is moral if it can be seen following a universal law' – law that can be accepted by all rational agents – and actor respects people such way they are not used only to some purpose, but are treated as ends themselves [14, 15].

Utilitarianism is an ethical viewpoint where the outcome of an action defines whether the act is ethical or not. It is commonly used in situations where there are different options to act. In those situations, the evaluation is made to find the act, which most brings the utility and thus is the most ethical by its consequences [14].

Whilst deontology is focusing on rules and utilitarian consequentialism to results of acts, the virtue ethics is focusing more on how to be good — to have a good character and embody the virtues in one's own person. Thus, virtue ethics is based on the character and virtues, which the person has internalized. A virtuous individual seeks to develop her own character such way that she nurtures the "good" characteristics and lives according to those. This means that one acts as to be just, honest, kind etc. as well as according to cardinal virtues, prudence, justice, temperance and courage by Plato (see [16, 17]).

Notable is that all of those three ethical approached have been criticized and have flaws. Nevertheless, it can be plausible argued that if some action seems to be unethical by all of those it most probably is unethical and *vice versa*. Thus, using all of those we can analyze with some certainty the ethical problems of usability design in the above tendering case.

4.1 Ethical Problems and Possible Solutions

The main problem of the usability design activities seemed to be the focus shift from users to requirements per se. The principles of human-interaction design emphasize the real use and users as the core units of analysis, but due to pressures to success in the forthcoming verification tests, the focus shift was seen necessary by the vendor. A direct implication was that some of the detailed usability redesign recommendations were also conflicting with general usability guidelines and best practices. From the deontological view, bypassing the universal laws of usability and user-centered design implies an unethical behavior.

However, IT vendors participating in tendering must truly believe that their systems fit for the intended purpose as otherwise they would not partake in tendering at all (unless their motives are unethical already in the beginning). What if the vendor knows, and it is true that their system fits very well to the real and intended purpose, yet which are not well formulated in the RFP, and thus the vendor needs first to comply with the RFP with the means discussed, in order to win the competition, and only later show how good their system is for the purpose. It seems to be acceptable from the utilitarian point of view to just try getting the test passed, in order to win the contract, if the system actually is good and could fulfil the needs of real use. On the other hand, vendors' are not able to compare their system against other systems (this is why the RFP exists), and know whether their outcome brings the most ethical outcome despite the possibly unethical acts. Moreover, for the procuring organization, the utility of the outcome would presumably be better, if the end-users were kept in focus all the time during usability design and whenever the systems were developed. Thus, the utility of the outcome is debatable. From the view point of virtue ethics (winning fairly) and deontology (way of winning cannot be seen as following a good universal law), the usability design acts of the case are not justified although vendors' could be seen as "pursuing good" from the utilitarian perspective, they are not winning fairly and thus not virtues[2]. As it seems not to be virtuous and good universal rule to develop systems and design usability only for the testing purposes and abandon making actual improvements for the end-users, we can state that usability design of the case was ethically problematic from all three perspectives.

When the usability design is distorted towards RFP requirements and performed distantly from users and use contexts (see the situation in Figure 1), it is hard to come up with the tendering process where the competitors could be sensibly or justly evaluated – in the sense that are they possibly the best provider for real needs of organizations. In that kind of situation the vendor has almost impossible to act in the tendering so that those acts could be seen ethical. Vendors usually do not have possibility to inform and suggest revision for the RFP. They have two possibilities: First, to try to fulfill the RFP requirements, which leads to unethical outcomes shown before. Other option is to withdraw from competition and let the other vendors to have that contract. The problem is that in both cases the outcome is not possibly based on real needs and will not be the fair one compared the situation where the RFP would be appropriate. Situation is not obviously fulfilling the requirements of deontology or virtue ethics. From the utilitarian basis, the outcome could be good, if the winner by some chance has the best solution for real needs — if there is a blessing in disguise.

The main reason for unethical usability design actions (breaking the rules, not being virtuous) was the rigid and detailed RFP created by the purchasing organization. In general, procurement policies aim at 1) acquiring the best technology solution 2) protecting integrity of the process and 3) ensuring equality of all bidders [18], where the rigid and detailed RFP can support at least the latter two. Strict and unalterable definitions for the testing try to ensure that every vendor is on the same line, each is fairly treated and has equal information. The challenges of a rigid and detailed RFP, as experienced in our case, are that a) the assumptions about users and work made during RFP creation must be valid and that b) there are no simple ways to alter the requirements, criteria or processes during the tendering.

Of course, the process can be preferred to be unalterable because there is possibility to end up situation of a complaining vortex. However, if that is the reason it seems that we are avoiding one problem with other one and with a worse outcome. In Finland, there is a possibility for the participants of the procuring process to appeal to Market Court (see [19]). Thus, the possibility to inform problems in the RFP could actually reduce the amount of appeals to Market Court, which means that problems could be solved during the process not after it — and this most likely

[2] Authors understand that companies cannot make moral choices because they are not rational agents. However, in this paper we treat vendors as rational agents, because there are people in companies who make those decisions, and it enables our arguments to be clearer.

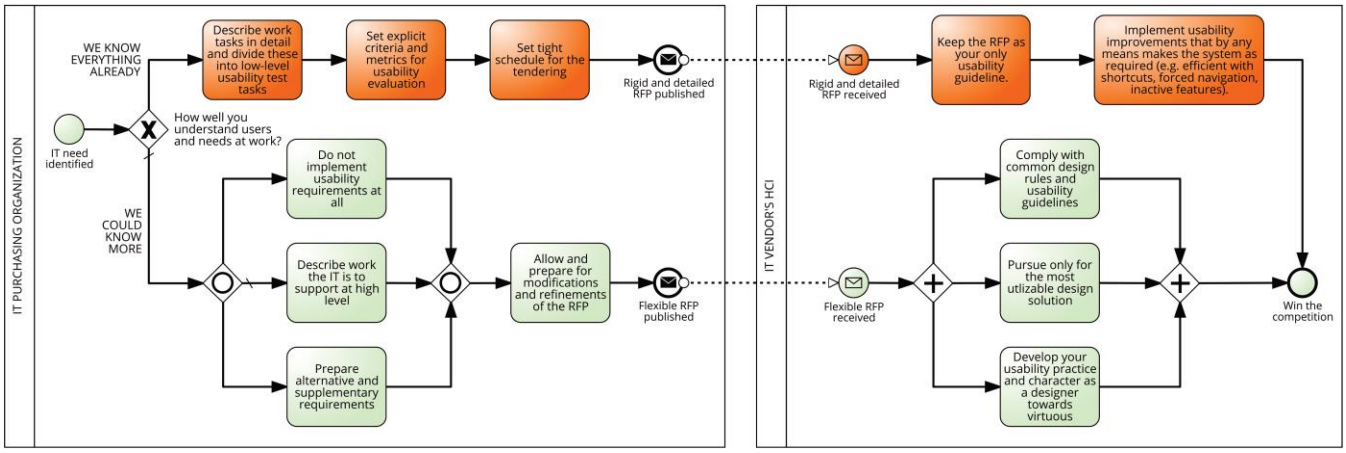


Figure 2. Two paths of RFP creation and subsequent unethical (dark colored) and ethical (light) usability design actions.

would be a faster and more rational way to handle the problems of RFP than Market Court. Ethically it seems justified to give this possibility, because it may encourage vendors to point out problems which come up with poor outcomes (utilitarian) and would give advantage to vendors who seek true benefits rather than who just try to win competition – more utilitarian, deontological and virtuous than the current situation.

On the other hand, if the requirements and procedures are not frozen by the RFP, the tendering and testing processes may prolong remarkably. If the RFP needs to remain unalterable, it must have characteristics that do not enforce unethical usability design actions i.e. to prevent designers from breaking the rules of general usability guidelines (e.g. heuristics by Nielsen [12]). Thus, not only the RFP needs to define what is needed, but also be careful in how that is needed. We find two extremes of characteristics appropriate for implementing ethical usability design: Either usability requirements and criteria are so abstract and generalized or so system-specific and tailored that unethical acts are not needed. On the basis of our case, this means that e.g. efficiency measured as a number of steps in a specific task is not the most appropriate formulation of a usability requirement. In literature, usability requirements that have alternatives [6] and requirements that are supplementary [4] have been suggested. The most abstract option is not to include usability requirements at all, if proper user performance based requirements cannot be defined [1]. We recommend generalizing the description of (work) tasks the system is to be support and concentrating on task outcomes when defining and evaluating usability requirements. For example, instead of asking to "send email to John", we would e.g. formulate that "the system must support communication between people" or "inform your colleague about issue X", which leaves space for the system implementation (virtue/deontology) but also increases the validity of the requirement (leading presumably better utility). The process model in Figure 2 represents this recommended path of RFP creation (below/light shaded) as well as the path of the case study (above/dark).

Other related solutions for the problems of current RFPs could be possibility to cease the contract with the winning vendor, if the real project turns to be too away from situation that was defined in RFP. That would also encourage vendors to aim at solutions, which are usable in reality instead of optimizing the system for testing purposes (assuming that the RFP is flexible enough as discussed above). In addition, there could be an independent foreman whose task is to evaluate is the IT project done properly and ethically. The foreman could then redirect or cease the project, if problems are not corrected [20]. Especially, the foreman would be appropriate in large projects where she has no position either on customer's or provider's side – the foreman just controls that everything is done properly like financial auditors are controlling financial issues in audits (see [20]). In large IT tendering projects for ensuring in the first place that there are not so much problems, the customers could have a pre-project, which aims to get an idea of real needs of users and organization in more deeply and thus come up with a proper RFP e.g. with the approach of Work Informatics (cf.[21]).

These aforementioned solutions would improve RFP and overall IT development projects by emphasizing the truthfulness (Deontology and Virtue ethics) and thus ensuring the better outcome of projects (Utiltarianism) and hence would serve an ethically responsible way to develop and procure information systems.

5. CONCLUSIONS

The ethical viewpoint contributes the IT research and practice and helps to understand the role of ICT for society [7]. Thus, ethical issues must be brought up to discuss the problems of IT tendering and RFPs. Our selected case experiences about usability design recommendations show how the detailed and rigid, yet ambiguous RFP with a tight schedule set the system usability designers in situation where they were tempted and even forced to make such usability proposals which look good – yet possibly unworkable – in order to avoid the loss of procurement. The situation is not ethically desirable because it persuades the IT provider towards proposal which is not the best (consequentialism) or truthful (e.g. Kantian deontology and Virtue ethics). Our claim is that the RFP and procurement processes should be altered such way that there is possibility for participants to suggest a change for it, if there is some crucial problem found. Alternatively, the RFP could be flexible enough by containing generalized or system-specific descriptions of user needs and usability requirements, which would not attract unethical design actions. Likewise, an option should be kept for exiting the project, if the real project does not meet the specifications presented in the RFP. That would encourage vendors to point out, if the RFP is failing to meet real issues or it is too vague. This way, vendors would be encouraged on investing in towards the actual performance of the system rather than trying to please the next evaluators.

6. REFERENCES

[1] Jokela, T., Laine, J., and Nieminen, M. Usability in RFP's: The Current Practice and Outline for the Future. *Human-Computer Interaction. Applications and Services*. Springer Berlin Heidelberg (2013) 101-106.

[2] ISO 9241-11:1998 *Guidance on usability*. International Organization for Standardization, ISO 9241-11 (1998), http://www.iso.org (1998)

[3] Winkler, I., and Buie, E. HCI challenges in government contracting: a CHI'95 workshop. *ACM SIGCHI Bulletin*, *27*(4), ACM (1995) 35-37.

[4] Carey, T. A usability requirements model for procurement life cycles. In Carey, J. M. (Ed.) *Human factors in information systems: An organizational perspective*, 2, Intellect Books (1991), 89-104.

[5] Kushniruk, A., Beuscart-Zéphir, M.C., Grzes, A., Borycki, E., Watbled, L.,and Kannry, J. Increasing the safety of healthcare information systems through improved procurement: toward a framework for selection of safe healthcare systems. *Healthcare quarterly (Toronto, Ont.), 13 (2010)*, 53-58.

[6] Lauesen, S. Usability requirements in a tender process. *Proc. Computer Human Interaction 1998.* IEEE (1998), 114-121.

[7] Thorén, C. Approaches for inclusion of usability and accessibility in ICT procurements. In *Proc. UITQ 2005*. (2005) 39-42.

[8] Nieminen, M., Laine, J., Teräs, S., Runonen, M., Kalakoski, V., Valtonen, T. and Boryzki, E. How to involve users in government system procurement? In *Proc. NordiCHI 2014*, ACM (2014) 805-808.

[9] Stahl, B. C., Eden, G., Jirotka, M., and Coeckelbergh, M. From computer ethics to responsible research and innovation in ICT: The transition of reference discourses informing ethics-related research in information systems. *Information & Management, 51*(6), (2014) 810-818.

[10] Baskerville, Richard L. Investigating information systems with action research. *Communications of the AIS* 2.3es (1999) 4.

[11] Tarkkanen, K. and Harkke V. Evaluation for Evaluation: Usability Work during Tendering Process. *CHI EA '15 Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM Press (2015) 2289-2294

[12] Nielsen, J. *Usability Engineering*. Academic Press,(1993).

[13] Dumas, J., Redish, J.: *A practical guide to usability testing* (revised ed.). Intellect Ltd, Exeter (1999).

[14] Feldman, F. *Introductory Ethics*. Prentice-Hall (1978).

[15] Kant, I. Originally *Grundlegung zur Metaphysic der Sitten*, Several translations, main translation: Liddel B. Kant on the *foundation of morality - a modern version of the Grundlegung*: Indiana University Press (1970).

[16] Frede, D. Plato's Ethics: An Overview, *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.),http://plato.stanford.edu/archives/fall2013/entries/plato-ethics.

[17] MacIntyre, A. C. *After Virtue: A Study in Moral Theory*, v. 92, University of Notre Dame Press (2007) 443-447

[18] Sieverding, M. Choice in Government Software Procurement: A Winning Strategy. *Journal of Public Procurement*, 8(1) (2008) 70-97.

[19] Heimo O., Koskinen J. Kainu V. and Kimppa K. (2014), Problem of Power: The Missing Agent. In Buchanan E., de Laat P., Tavani H. and Klucharich J.(Eds.), Ambiguous technologies: Philosophical issues, Practical solutions, Human nature: *Proceedings of the tenth International conference on Computer ethics – philosophical Enquiry* (2013) 160-169.

[20] Kainu V. and Koskinen J. Why (an) Ethics code for information system development needs institutional support: there is even an upside for computing practitioners and businesses. *Ethicomp 2014*. (2014)

[21] Nurminen, M. Work Informatics – An Operationalisation of Social Informatics. In J. Berleur, M. Nurminen and J. Impagliazzo (Eds.), *Social Informatics: An Information Society for all? In Remembrance of Rob Kling* Vol. 223, Springer (2006), 407-416.

KTP and RRI – The Perfect Match

David Kreps
University Of Salford
The Crescent
Salford
d.g.kreps@salford.ac.uk

Jessica Blaynee
University Of Salford
The Crescent
Salford
j.blaynee@edu.salford.ac.uk

Maria Kutar
University Of Salford
The Crescent
Salford
m.kutar@salford.ac.uk

Marie Griffiths
University Of Salford
The Crescent
Salford
m.griffiths@salford.ac.uk

ABSTRACT

Businesses working with universities are in an optimal position to overcome perceived barriers to the uptake of Responsible Research and Innovation (RRI) for Business and Industry. This paper sets out case study evidence within a particular framework for such university-industry partnerships, to support this assertion, and suggests that the framework in question could with minimal development become an even better vehicle for encouraging such uptake, and an example for other EU countries to follow.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**] Ethics, Regulation

General Terms

Management, Design, Economics, Human Factors.

Keywords

RRI, Innovation, Research Ethics, KTP

1. INTRODUCTION

Businesses working with universities are in an optimal position to overcome perceived barriers to the uptake of Responsible Research and Innovation (RRI) for Business and Industry. This paper sets out case study evidence within a particular framework for such university-industry partnerships, to support this assertion, and suggests that the framework in question could with minimal development become an even better vehicle for encouraging such uptake, and an example for other EU countries to follow.

2. KNOWLEDGE TRANSFER PARTNERSHIPS

Knowledge Transfer Partnerships (KTPs) are billed as Europe's leading transfer mechanism, designed to help companies increase their competitiveness and productivity through the improved use of the technology, skills and knowledge that resides within UK academic institutions - the knowledge base. The framework, which is UK government funded, is a 40-year-old scheme that facilitates innovation and productivity in UK businesses, by linking them with academics in universities who are experts in their field. The partnerships include three key stakeholders; the academics, the company and the associate, who is typically a recent under- or postgraduate. The process is facilitated by a KTP advisor who facilitates the development of the KTP from bottom up and then who guides the KTP through the approval process. This official advises and mentors all stakeholders throughout the lifetime of the project. The lifetime of a KTP can be anything from six months to thirty-six months and if the problem or project dictates, the company can have more than one KTP. There are many benefits that a KTP presents to the three stakeholders: (i) firstly the academics have a unique opportunity to apply research to a real world situation, and identify novel research themes that often emerge through the KTP. Importantly there is opportunity to apply their expertise and knowledge to solve real-world business problems, and then these situations can be used to inform and develop relevant teaching materials and written up as case studies. A major feature of the KTP is that it should be a challenge for the academics, so that knowledge is transferred in this direction as well. There is an expectation that papers related to the project will be published and the funding will contribute to the regular Research Excellence Framework assessments of UK academic output. (ii) The Associate (two of the authors are past and present associates) has many opportunities in the role of project owner, as well as the wealth of experience that can be gained from managing a distinct project. The role often requires that they work closely with senior management in the business. The KTP project is typically a catalyst for a change management process. Furthermore the KTP 'package' includes the opportunity to study for a higher degree and has a significant personal training budget, ensuring there is a continuous learning and knowledge exchange. (iii) The Company's benefits are unique too: they have the support from the academic team, and from the KTP advisor over the lifetime of the project; and other student projects are encouraged so there is an immediate relationship building with the University involved. Typically a company undertakes a KTP to enhance an existing project, process, to develop or implement systems or a strategic direction. But ultimately a KTP is most often (though not always) undertaken to improve efficiency, embed competiveness and to generate wealth.

The format has changed little over the last forty years. In 1975 the Department of Trade and Industry (DTI) launched the Teaching Company Scheme (TCS), which became the KTP scheme in 2003, and the Technology Strategy Board has been the main funder from inception until the recent change of name to Innovate UK. What hasn't altered in the past 40 years is the structure and main focus of the KTP - because it works. In 2014 there were 642 active partnerships in the UK, with 105 different knowledge base partners involved; 10% at micro companies, 39% at small companies, 25% medium companies and 26% at large companies. This gives an indication that there is a natural spread across company sizes demonstrating that the size of the company is not a limiting factor of the KTP. Recent statistics from the company report [1] present an overview of the share of KTPs across the UK: Queen's University of Belfast have a staggering 31, University of Newcastle 31, University of Strathclyde, University of Nottingham, and University of Manchester all with 23 KTP's. Beyond this is a more even spread of KTPs across the UK. The University of Salford ranks highly with successful KTPs in the North West and had 12 KTPs running at that time period. Companies typically contribute 40-50% of the cost of hiring an Associate (who remains an employee of the university during the project whilst working on the company's premises), for a two year discrete project, supervised by a team of academics and company members. The UK government provides the balance of the funds.

We argue that this structure is an ideal vehicle through which the aims of RRI can be fostered. The authors have experience as both KTP Associates – placed in businesses for two-year research and innovation projects - and Supervisors, creating and managing such projects from the University side.

3. RESPONSIBLE RESEARCH AND INNOVATION

The development of the notion of RRI may be traced from 'Science and Society' to 'Science in Society' and on from there to Responsible Research and Innovation. This paper is concerned with a particular focus on the relationship of RRI with Business and Industry. Nonetheless, there are aspects of the wider literature on responsible research that are of significant importance to any field in which the notion of RRI applies. As Stilgoe points out, "Science has been conventionally invoked by policy as emancipatory. This has allowed scientists and innovators considerable freedom from political accountability." Where scientific research and innovation, in other words, has resulted in contrary outcomes, a more responsible approach may have made a substantial difference. 'Science *and* Society', in short, presumes an unquestioned narrative in which all scientific research and innovation is considered distinct from, and applied to, society, as an assumed 'good' that improves the lot of all. The field of Science and Technology Studies (STS), however, situates Science *in* Society, has suggested over the past several decades that "science and technology are not only technically but also socially and politically constituted," [2] and neither distinct from, nor simply applied to, society, and thereby not necessarily 'good' for all whom it touches.

A new focus upon responsibility with regard to scientific innovation, promoted by such studies, and by a range of high-profile public controversies - from medical innovations such as thalidomide to agricultural innovations such as genetically modified organisms - has failed to halt the number and scale of instances where the continued application of science *to* society has resulted in less than optimal outcomes. Relying, it would seem, on a legal and policy framework of accountability, liability and evidence concerning the 'products' of science and innovation, as the levers by which to enforce responsibility, has proven ineffective. The retrospective nature of such approaches makes them - in practice - very difficult both to reliably trace, and then effectively enforce, especially considering the often complex nature of overlapping and interdependent innovations. STS notions such as Hardin's 'tragedy of the commons,' [3] where the common good is harmed by the common sense behaviours of many individuals, e.g. the overconsumption of resources, and Winner's [4] notion of the need for representation in innovation of the interests of those whom it might affect, have together with many other scholarly approaches encouraged a more forward-looking approach to science and innovation governance.

One outcome of these developments has been the recent publication of the Rome Declaration on RRI, which seeks to promote a new settlement for research and innovation that could be described as 'Science *with and for* Society'. Responsible Research and Innovation (RRI) in this context is a multidimensional concept that includes six key issues: public engagement, gender equality, open access, science education, ethics, and governance. The Horizon2020 call for proposals soliciting funding bids in this arena describes RRI as an attempt: "To allow all social actors (researchers, citizens, policy makers, business, third sector organisations, etc.) to work together during the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of European society." [5]

4. KTPS AND RRI

The overlap between the aims of the Knowledge Transfer Partnership scheme and the core messages of the Rome Declaration on RRI is such that many of the aims of RRI are indeed implicit in the scheme, and could – with little substantial change – be made much more explicit.

KTPs enhance engagement by bringing otherwise disparate actors – universities and private companies – together; KTPs improve gender equality through the university appointment process; KTPs enhance science education through bringing scientific researchers into the work-place to transfer knowledge and expertise; KTPs disseminate best practice in ethical working from the academics to the private sector; the very nature of such Knowledge Transfer Partnerships is to open access to publicly funded research; and the structure of the KTP is an aspect of public policy governance that is very much in keeping with the aims of RRI.

The nature of the KTP, with an associate employed by the university, and significant input from the academic team, means that KTPs may become drawn in to the research governance procedures of the university. Research governance has evolved considerably over the lifetime of KTPs; during this time research in universities in many countries, including the UK, has moved from being a largely unregulated activity to one which is increasingly formalised and drawn into institutional processes which seek to regulate research activity [6]. The rise of the Research Ethics Committee (and its North American counterpart, the Institutional Review Board) with a remit encompassing the full range of disciplines in the institution, may be viewed as an institutional response to encourage RRI, although the practice is not without criticism [7] and it is clear that whilst institutional oversight is increasing, the governance mechanisms employed in these practices have not yet evolved to become fully mature, in particular with regard to openness, transparency and review of the

governance mechanisms themselves. University research governance policies and processes, together with those of the research funding organisations are concerned with the core concepts of harm, and its avoidance, risk, informed consent and anonymity of participants. The 'researchers' and 'research' that fall under the research governance umbrella may vary but typically this is defined to include research with human or animal subjects, or which may impact upon the institution's reputation. Institutional policies which require for research carried out within a KTP to be subject to research ethics review vary but this is increasingly likely, and may be seen as a contributory process to the effective, adaptable and responsive oversight of research which forms a core element of RRI [8]. There is, however an inherent tension within the KTP whereby the commercial partner is often unused to the bureaucratic and regulatory constraints of research in the University environment, and there may be expectations of a greater agility than such systems allow. Despite this tension, the research governance environment may act as a supportive tool to embed RRI in the KTP process.

We believe, moreover, that with enhancement, this match between KTP and RRI could be improved still further.

5. EVIDENCE

We now present some examples to support this argument, set within the context of our experience of a currently running KTP, on which all four authors are engaged. Firstly, as a Case Study, we demonstrate how within the early months of the current partnership, the relationship has led to immediate improvements in RRI in the business, as well as encouraging the uptake of the UK's Digital Inclusion Charter - something which extends the benefits originally envisaged in the creation of this particular partnership. Secondly, with reference to a project the business in question is engaged with, digital training for over 65s, the paper examines how potential barriers to RRI in the host company have been identified and overcome. These examples we set in the context of the more general digital inclusion literature, with which two of the authors are very familiar. These help to illustrate the challenges that many businesses face when trying to balance their immediate business needs against the longer term benefits of research informed development. The tensions arising from these challenges can act a barrier to the uptake of RRI, whereby the perceived constraints of RRI are seen as an additional cost.

5.1 Updating Company X's Ethical Procedures

The first example concerns a currently running KTP with Company X. At the beginning of the KTP one of the objectives was to draw up a full Ethical Approval Application for the project, to be put through the University of Salford Ethical Approval process. To do this the associate began to compile information on existing ISO standards and discuss with the academic members of the team what constituted academically sound Informed Consent and Data Protection procedures.

From this the following things were recommended: A redraft of the Informed Consent form, stipulating (i) Data for research purposes and third parties of a research nature (not just marketing) (ii) Data will be anonymised *(apart from IC for video/ voice use – where it can't be anonymised)* (iii) That participants can contact Company X to have the data removed from their database (iv) The length of the project, and how long the data will be in active circulation before archiving. This would also involve (a) a restructure of the company's data folders and server access so that all Personal Data captured by the company can only be accessed by those that require the data for their role. (b) Plans to update the ICO notification to include owning/having data for research purposes, and (c) Clarification that all work with Big Data requires informed consent for Third Party use when clients to ask Company X to use it.

The new informed consent forms were trialled during a project with Client A, undertaken by the company, in connection with the KTP (see below) and are now in full use at Company X across all projects. There is a now a separate research folder which only certain team members have access to and the ICO will be updated at renewal in mid 2015.

5.2 Digital Inclusion and Aging Population Study

The research element of the KTP required a distinct mini-research project to look at a specific area where user experience could be improved. Company X has a strong interest in accessibility and digital inclusion, and it made sense to take this further.

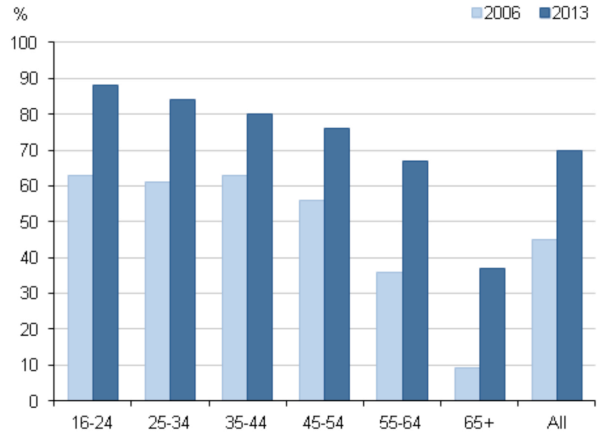
One suggested area of focus for the KTP was the UK's ageing population. 10 million people in the UK are over 65 years old. The latest projections are for 5.5 million more elderly people in 20 years' time and the number will have nearly doubled to around 19 million by 2050. Within this total, the number of very old people grows even faster. There are currently three million people aged more than 80 years and this number is projected to almost double by 2030, reaching 8 million by 2050.

The KTP proposed to investigate how older users interact with digital services and technology, also assessing the needs of disabled user groups, to understand how physical and cognitive disabilities should inform digital services.

Figures show that since 2006 there has been a sizable increase in daily computer use for adults aged 65 [9]. This has led to senior users becoming the fastest growing demographic on the Internet.

Daily computer use by age group, 2006 and 2013

Figure 1: Daily computer use by age group, 2006 and 2013



Source: Office for National Statistics

[10]

At a high level, the KTP team are keen to explore the challenges and opportunities that will arise from initiatives such as the Government's Digital by Default strategy. Researching how different user groups engage with online services and brands across devices, in the home, at work and at play. As society becomes more web-enabled User Experience will be more important for social scientific analysis, and ensuring this demographic can engage with these services will be key to their success.

By developing our understanding of the goals, challenges and motivations of this user group Company X will be able to provide relevant advice and guidance to aid organisations in how best they can meet the needs and expectations of this increasingly important user group. Moreover, as part of the Digital By Default scheme there has been a growing obligation for companies who take up digital provision to also incorporate the Digital Inclusion Charter, [11] an agreement for a cross-sector partnership, to not just scale solutions but to ensure they're fit for purpose and reduce the number of people who are offline by 25%. These two agreements appear to need to work together in order to truly be effective.

Some of the stipulations of the Digital Inclusion charter - a common definition of basic online skills and capabilities, and the need for a shared language - are already familiar in a private sector environment such as UX, which often preaches simplicity. But when digital inclusion is understood as a incorporating web accessibility, Company X – already strong in the latter – are in a very good position to be work with and feel a part of this charter. Additionally Company X frequently do 'knowledge transfers' with their clients, and perform particularly thorough handovers in order to pass on digital knowledge and best practices to their partners. This aligns with the charter's aims of "embedding digital inclusion into partners' communications activity to encourage people, small and medium enterprises (SMEs) and voluntary, community and social enterprises (VCSEs) to take the first steps to going online". [11] (As shown with a recent project Diversity Role Models.)

5.3 The Digital Inclusion Project (Client A)

With the introduction in the UK of a new social security system, called Universal Credit, which will require welfare recipients to manage their help online, Client A, a Housing Association, have taken on a dedicated digital inclusion officer. From October 2015 to March 2015 Company X hosted, and then ran the digital inclusion sessions using the Tinder Foundation's Learn My Way programme. [12]

During sessions participants were separated into groups by what they wanted to learn. There were groups of up to three, more typically one-on-one. One-on-one engagement is traditionally used in usability studies and this allowed for a greater amount of observation and for participants to voice their queries and interests as they occurred. Real time observations and reactions were key to this type of research.

Each session was broken into 2 weeks. The first week was an open exploratory session, providing people with the opportunity to express their problems and bring their own devices. The second week was structured learning through Tinder Foundation's Learn My Way programme. There were no research 'groups' *per se* but participants could be retroactively designated to categories such as their need to learn digital skills; imperative *vs* non-imperative; device ownership, and their existing experience: from none to low.

The primary research goals were to explore what users had the most difficulty using online and how many tasks were done digitally on a regular basis. Our hypothesis at the beginning of the study was that those with virtually no experience would have no social media awareness but have some experience of email; and those with more experience would be 'online' in the sense of shopping, with a possibility of being on social media, too. The secondary research goals were to try our new informed consent form and ethics measures detailed above, and also for Company X to incorporate the findings, regarding what users had difficulty using, into their own proposals and projects to improve their solutions. By doing this they would further be incorporating the Digital Inclusion charter into their practices.

Part of Client A's goals was to assess engagement with digital courses, as this would demonstrate impact of the course on motivation. The word 'engagement' is also typical of the type of language used in the discourse surrounding digital inclusion as a theme. Although understood to be a broad term, for digital, 'engagement' incorporates so much in people's lives it could not be broader. It was at this point that a pattern in behaviour became clear between those who *had* to learn (imperative) for either job hunting/employment, paying bills, and social absolution, on the one hand, and those who *wanted* to learn (non-imperative), whose motivations ranged from curiosity to specific creative hobbies such as photography. Both categories are 'engaged,' but have a complex relationship with the computers and devices, (making it an angled scale) and suggesting engagement is temporal, something to be cultivated, rather than simply 'switched on'.

For example, it was found that those already quite confident with their devices, were oblivious to many features, or not using them correctly. This would in fact suggest a lack of engagement. In the same vein, it was possible for people to be highly engaged in some aspects of the web but completely disengaged with others. The biggest example of this was money and financial transactions. Many people who shopped online would not bank online, or use Paypal (and as a result could not use eBay, a benchmark of comfortable online shopping.) Those that were happy with their devices and only brought them in to do the course on, were found to be struggling with the maintenance of them; backing up photos one by one because they were unable to sync, losing application access because the update message was unclear ("Please update your browser? This totally gets me. Don't understand a word"[9]). Security and Privacy was also a key issue which inhibited engagement, but not only in the traditional way of keeping one's details safe, but also with software and programmes "tracking" them ("I just signed into Gmail now my name is all over the place everywhere here"[9]). These issues and many others like it could be broadly condensed into a problem with information transience, where users are unsure where online information exists or comes from and are unable to differentiate between their computer and the Cloud, or between Wifi and broadband. In the most extreme cases this meant that every online service became a perceived threat to their real world goals. For example, in once case, a job centre emailed a participant's CV to her husband's email address, making him hostile towards her because he believed this meant that the job centre had somehow hacked his computer.[9]

Only when users had grasped where information was stored online could they take control of their security online. Many who had anti-virus protection in place were still, however, periodically victim to viruses and hacking by clicking pop-ups which appeared

to be genuine system messages due to their design. The effect of this was that they then began to distrust the system messages and either stopped using the web altogether or refused to make any financial transactions. An irony can be drawn here in that it is these messages and pop ups which make the largest effort to contribute towards 'a shared language'. The same could be applied to spam or junk emails and there was a serious negative impact on users who regularly received emails from companies without knowing why. The approach to the Cookie policy was also damaging; far from being informative the notification messages were often barriers to further use, and there remained some confusion arising from a perception that if the cookie message was prominent, it probably meant more data was going to be used 'badly'. This meant that sites that didn't display a cookie message or kept it low key were perversely deemed to be safer. Security and issues around data and privacy concerns are the greatest challenge when encouraging or discouraging user interaction with specific services (banking, home/personal services) and further KTP research could explore whether further legislation is necessary, e.g a further examination of the cookie policy.

In sum, the experience within Company X of the KTP brought about a new and much deeper engagement with the concerns of RRI than might otherwise, in the normal course of their commercial activity, have occurred. The ethical processes of the company received a thorough renewal, informed by the processes at the University. The already admirable interest and expertise in Web Accessibility expanded into a fuller engagement with Digital Inclusion as a whole through the Digital Inclusion Charter, immediately of value in work with the P&P Housing Association and commercial research into the digital user experience of over 65s.

6. CONCLUSION

We conclude that, with the benefit of lessons learned from the case studies, the KTP scheme could include more explicit stimulus towards RRI, making KTP-style university-industry engagement an ideal platform for expanding RRI, and an exemplar for other nation-states, or indeed the European Commission, to emulate, in their funding structures. Our analysis has considered the barriers and enablers for RRI in university-industry partnerships generally, such that this may inform the design of such partnerships elsewhere.

Our recommendation, within the UK, is that aspects of the RRI process should, we believe, be embedded into the KTP application process when the proposal is written and the workplan is scoped out, and also in the evaluation process once the project is complete. This offers an ideal platform to firmly introduce/embed RRI practices into a business and transfer that knowledge, leaving an RRI legacy. A similar incorporation of RRI into other schemes, in other countries, is also recommended.

7. References

[1] KTP 2014 Company Report. *A Study on Video Browsing Strategies*. Technical Report. University of Maryland at College Park.

[2] Stilgoe, J., Owen, R., and Macnaghten, P. 2013. Developing a framework for responsible innovation. *Research Policy.* 42, 9, (Nov. 2013), 1568-1580. DOI= http:// doi:10.1016/j.respol.2013.05.008.

[3] Hardin, G (1968). "The Tragedy of the Commons". *Science* **162** (3859): 1243–1248.

[4] Winner, L., 1986. The Whale and the Reactor: A Search for Limits in an Age of High Technology. University of Chicago Press, Chicago.

[5] http://ec.europa.eu/research/swafs/index.cfm?pg=funding

[6] Shaw, S., Boynton, P. M., & Greenhalgh, T. (2005). Research governance: where did it come from, what does it mean?. *Journal of the Royal Society of Medicine*, *98*(11), 496-502

[7] Haggerty, K. D. (2004). Ethics creep: Governing social science research in the name of ethics. *Qualitative sociology*, *27*(4), 391-414.

[8] Sutcliffe, H., & Director, M. A. T. T. E. R. (2011). A report on Responsible Research and Innovation. *European Commission, Brussels, Belgium*.

[9] Interview by J. Blaynee.

[10] ONS 2013. Internet Access - Households and Individuals, *2013*. http://www.ons.gov.uk/ons/dcp171778_322713.pdf

[11] https://www.gov.uk/government/publications/government-digital-inclusion-strategy/uk-digital-inclusion-charter

[12] Tinder Foundation's Learn My Way programme

Who Is To Change? Nudging and Provocative Communication Discussed through Løgstrup's Ontological Ethics

Thomas Dyrmann Winkel
Aalborg University
Teglgaards Plads 1, 11
9000 Aalborg
winkel@hum.aau.dk

Thessa Jensen
Aalborg University
Teglgaards Plads 1, 11
9000 Aalborg
thessa@hum.aau.dk

Søren Bolvig Poulsen
Aalborg University
Teglgaards Plads 1, 11
9000 Aalborg
bolvig@hum.aau.dk

ABSTRACT
This paper discusses nudging and provocative communication as possible approaches to designing behavioural change concerning minimisation of waste within the framework of Løgstrup's ontological ethics. Waste management companies are confronted with ethical concerns as their course of action consequently affects their relationship with the citizens whose waste they manage. Waste management companies might be experts within their field, but they are challenged when entering new contexts and must therefore redefine or reframe their role in society. This became evident during an action research project as an ethical challenge was identified through a strategic workshop facilitated for AVV in relation to the Nulskrald project. The main focus of Nulskrald is citizen empowerment as well as organisational learning and responsibility. Through Løgstrup's ontological ethics the ethical demand, as it is posed by 'the other person' towards the 'I', will show concerns and possibilities for engaging citizens, while at the same time resulting in organisational development. Therefore, the research question is: what ethical issues and organisational implications exist concerning the use of nudging and provocative communication, respectively?

Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human factors, Human Information Processing

H.5.2. [**User Interfaces**]: User-centered Design

J.4 [**Social and Behavioral Sciences**]: Psychology, Sociology

K.4.1 [**Public Policy Issues**]: Ethics

General Terms
Design, Human Factors, Theory, Management

Keywords
Communication, nudging, ethics, RRI, design, critical design, behaviour change, provocative communication, Løgstrup

ETHICOMP 2015, September 7–9, 2015, Leicester, UK.

INTRODUCTION
The generation of municipal solid waste is a global environmental challenge as it leads to greenhouse gas emissions and ecological degradation [51, 32, 44]. Waste prevention is essential to ensure a sustainable handling of waste and a sustainable future. OECD defines waste prevention as strict avoidance as in not generating waste in the first place, minimising the use of dangerous substances, and increasing product reuse [31]. Waste management companies are established by the government to sustainably manage waste with respect to defined requirements [4]. These companies' task is to distance and separate waste from the societies producing it.

Some waste management companies also explore initiatives that go beyond the requirements established by the government to pursue their own visions of reaching the higher levels of the waste hierarchy constituting the default guideline to waste management: waste prevention and minimisation, reuse, recycling, and disposal [54, 27]. In doing so, the waste management companies are confronted with ethical concerns as their course of action consequently affects their relationship with the citizens whose waste they manage. Waste management companies might be experts within their field, but they are challenged when entering new contexts and hereby redefine or reframe their role in society. This problem became evident in this case study of a local waste management company.

The publicly owned waste management company AVV (Affaldsselskabet Vendsyssel Vest), located in Northern Jutland, Denmark, has a vision not only to collecting and managing, but also minimising waste [54]. They launched the Nulskrald (Zero-waste) project in 2013 to explore these opportunities through interventions within their area of responsibility. The overall ambition of the project is to minimise waste through consumer empowerment and behavioural change, and to redefine AVV's relationship with its citizens. Nulskrald has already gone through two phases and a strategic workshop was initiated to develop the directions of the third phase, hereby also which role to take on in relation to the citizens.

During the workshop the overall purpose of project Nulskrald was framed and defined as *Think the environment first*, which was expressed into three missions; *Start a movement*, *Motivate to minimise waste* and *Transparent business*. Three core areas were developed, and since then nudging and critical design through provocative communication has been included in AVV's

approach to the citizens. However, the two design approaches involve new perspectives on both citizens and AVV as an organisation itself. With this new role in society ethical issues must be addressed, especially as the role implies changing citizens' behaviour. This raised the research question: what ethical issues and organisational implications exist concerning the use of nudging and provocative communication, respectively?

RESEARCH METHODOLOGY

The framing of this paper departure from ontological ethics by Løgstrup seen in a design perspective discussing the use of nudging and provocative communication as a variation of critical design. The two approaches to change citizen's behaviour were produced during a workshop, which was part of an action research project. The two design approaches are discussed within the framework of Løgstrup's ontological ethics, focusing on the role of AVV and its different relationship with citizens due to the new role the two approaches entail. Finally, the paper ends with conclusive remarks on the ethical characteristics and organisational challenges posed by nudging and provocative communication.

BACKDROP OF THE PROBLEM

The workshop mentioned above was facilitated by two of the authors of this paper of which one is partly hired by the AAV. From prior interviews a need to establish a shared perception of the context in which they operate was recognised. This was necessary before they could collectively define which position to take and role to play in the third phase of the Nulskrald project. A wide selection of participants within the organization was invited – from the top management to those responsible for actually carrying out the work on a daily basis. The CEO, the Innovation Manager, the project manager, the internal communicator, one of the employed industrial PhD-fellows and the business developer participated in the workshop. During the workshop the facilitators took an active part in the discussion by questioning and challenging the participants to collectively develop the new insights and the future direction in their context, containing a high level of complexity.

A designerly approach was applied to embrace the complexity of navigating in a political landscape with many actors and numerous regulations. The six hour workshop was grounded in working visually by both mapping the existing state and negotiating the future and desired state. Two frameworks, the Actantial model [12] and the Strategic Pyramid [22], were selected to support the overall purpose of the workshop. The Actantial model is, traditionally, applied to theoretically analyse actions in works of literature due to its ability to decompose narratives into actants and hereby illustrate how they interfere and affect each other on the axis of *power* and *desire* to describe the influence on the axis of *transmission* [12]. Object Theatre [46] was applied to the Actantial model with the purpose of facilitating a rich dialogue among the participants as it can support the articulation and physical configuration of meaningful stories of professional practices of any sort [45], which has also proven to be valuable in design settings [36]. In practice, the different actants was characterized by picking artefacts from a broad and randomly picked collection; for instance a baby figure, a dollar note, a warrior, a house, a boat etc. The gained insights from this activity were recorded in the Strategic Pyramid, which has three levels ranging from the top with the overall purpose through vision and mission to the actual goals at the bottom. The reader is referred to Winkel et al. [54] for further discussion of the methods mentioned here.

Based on existing and new ideas, three core areas were developed to a conceptual level:

1. Increasing reuse of clothing, electronic products and construction waste by providing a better service in the associated shops and by re-designing the clothes
2. Encouraging citizens to unsubscribe the weekly deliverance of paper commercials through nudging by providing pre-filled unsubscribing forms
3. Promoting citizens to reflect upon food waste through critical design or provocative initiatives.

It should be noted that 'provocative initiatives' was later renamed 'provocative communication'. These three short-term goals are a shift of attention from recycling to reuse and minimisation of waste. This implies a new role for AVV to enter in relation to their citizens, which raises an ethical implication: how would AVV's relationship to the citizens be affected when these initiatives were implemented to achieve the desired transformation of the citizen's behaviour?

THE ETHICAL CHALLENGE

Two dominant directions were found during the workshop – these, as mentioned before, were nudging and provocative communication. Both can be perceived as approaches to design – yet in different manners and with very different approaches. These two conceptually different approaches to design raise the question of which role designers should take with respect to how their designs influence people and the world.

"To think of designed things and design actions as material articulations tell us that design should be considered as a decision and direction embodied in all things human bring into being. Design is conditioned by its orientations, directions and capacities, while at the same time conditioning human beings, things and the world. Design articulates possible conditions through materialities." [21].

With this in mind, ontological ethics as explained by Løgstrup can sensitise the designer to the problems, challenges and opportunities posed by the two different design approaches.

FRAMING OF THE PROBLEM

1.1 Løgstrup's Ethical Demand

Ontological ethics take their starting point in the dyadic meeting between two people, the 'I' and 'the other person'. The ethical demand arises from 'the other person', who meets the 'I' with an unspoken plea for mercy, respect, and trust. These are some of the key concepts, also called life manifestations or expressions of life [25, 26], in Løgstrup's ontological ethics [24]. His setting is the meeting, in which the ethical demand, being unspoken, has to be acknowledged nonetheless. The 'I', in our case AVV, has to acknowledge the lifeworld, challenges, and problems, 'the other person' – in our case the citizen – faces in his/her daily life.

Løgstrup places the responsibility for bringing the ethical aspect of the meeting to life solely on the 'I'. Thus, AVV has a responsibility for any interaction with any citizen to acknowledge and recognise said citizen's lifeworld. Løgstrup [25] explains, how the ethical setting given by the life manifestations and acknowledgement of 'the other person', widens the scope for actions and choices, open to 'the other person'. Instead of just reacting, expressions of life enable the 'I' to actively change a

situation, both for the 'I' and 'the other person'. Engaging with the citizens and acknowledging their input might be one way of developing a movement or at least become part of the participatory culture, which exists on the Internet [16].

Løgstrup's stance on ethics is further developed by Pahuus [33], who focuses on the pedagogical implications of developing life manifestations. He points out, as does Løgstrup [23], how life manifestations either are present or not. Contrary to needs, especially artificially created needs in the western society of over-abundance, which can be fulfilled and created, life manifestations are present in every human being, but can be repressed or turned into destructive life forces. Thus, we meet another person with trust, a positive life manifestation, or mistrust, a negative life force. Trust enables, while mistrust destroys and hinders co-existence [26, 33].

With this in mind, a meeting between AVV and any citizen will always have other purposes than just the well-being and development of the citizen. Yet, as shown by Jensen [17], social media and participatory culture facilitate the development of life manifestations and learning, despite having other ends in mind as well. In the following, nudging and provocative communication as a type of critical design will be reviewed and discussed.

THEORETICAL OUTLINE

2.1 Nudging

Human's decision making is affected by cognitive shortcomings stemming from two kinds of thinking [e.g. 49, 19]. In short, dual-process theory describes the workings of these two kinds of thinking as two ways of processing information (for comprehensive collections of dual-process models, see [2]). Stanovich and West [43] conceptualises the different generic properties of the two processes as systems, which the authors label System 1 and System 2. Thaler and Sunstein [49] refers to the systems as The Automatic System and The Reflective System, respectively. Both systems are used when processing information and making decisions; however, they differ in the manner they do it. System 1 is intuitive, automatic, fast, and largely unconscious [43]. Thus, the system operates effortlessly. On the other hand, System 2 contrasts System 1 by operating in a controlled, slow and effortful way as it processes information consciously [45]. These systems are comparable to Piaget's [34, 35] concepts of assimilation and accommodation. Assimilating information means fitting it into existing categories without changing them, whereas accommodating information refers to confronting and adapting the existing categories or indeed creating new ones, thus modifying current knowledge. This implies conscious processing of the information and constitutes a learning process.

Nudging relies on this dual view of thinking in which the interplay between System 1 and System 2 results in systematic errors [49]. Both systems operate in parallel [10] and "[…] the accuracy of the (System 2) decision rule rests on the validity of a System 1 computation […]" [10]. Choices made in complex situations and under uncertainty, thus, result in the systematic errors [50]. As explained below, nudges are interventions designed to help citizens avoid these conceived errors by designing a context that presents a choice in a paternalistic libertarian way; that is re-arranging the choice set in a liberty-preserving way. But who chooses how a choice should be presented? Who has the power to influence others' lives legitimately? Do the ends justify the means? It is our contention that the way in which a choice is presented – that is, designed – must be determined by the ethical relationship between the two parties: the nudger and the nudgee, or the designer and the citizen.

The legitimacy of the choices made, can be challenged in a Løgstrupian sense, since it poses the opposite of enabling the citizen, in opening the possibilities and potentials hidden in every human being. Even if Løgstrup's ontological ethics can be seen as paternalistic, since the 'I' holds a small part of 'the other person' in their hands and is responsible for the empathetic treatment of the other human being [24], Løgstrup emphasises the opposite: "Thinking and imagination become equally superfluous. Everything can be carried out quite mechanically; all that is needed is a purely technical calculation. There is no trace of the thinking and imagination which are triggered only by uncertainty and doubt." [24]. The 'I' has to constantly doubt themselves to ensure the empathetic meeting of 'the other person'. As Thaler and Sunstein will show, the choice-architect – the designer of the given nudge – should not be bothered by doubt or uncertainty.

Thaler and Sunstein [49] describes the central idea of the term 'nudge' as a design approach – a choice architecture – that steers people's behaviour in directions that will improve their lives [49, 19]. Specifically, the intention of nudging is to make people's lives better "[…] *as judged by themselves*." [49, 47 [their emphasis]]. Nudging is understood as "[…] any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives." [49]. The theory behind nudging builds on psychology and sociology dating back over a century; however, recent insights from behavioural economics and social psychology underpin it to explain people's irrational behaviour. Individuals make bad decisions since they do not have unlimited cognitive abilities, complete information and complete self-control. In other words, people have a bounded rationality [41, 42]. Additionally, Thaler and Sunstein [48, 49] grounds nudging on 'libertarian paternalism', which should be the default design guideline behind every nudge.

Central to the concept of nudging is that it is not a mandate to act, because the approach to the design of a nudge is based on libertarian paternalism (see [49, 20, 38), a seemingly contradictory term previously defined by the same authors (50). The libertarian aspect refers to allowing people to maintain their freedom of choice in a choice context. The paternalistic aspect is when a policy "[…] is selected with the goal of influencing the choices of affected parties in a way that will make those parties better off." [48]. However, there exist different varieties of paternalism [47]. Some varieties only affect the means by which people try to reach their ends, while other varieties attempt to affect people's choices of ends. In addition, some varieties are 'hard', exercising a high degree of power. Conly [3] calls this coercive paternalism and claims that denying citizens choices may be liberating, allowing them to focus on "[…] the decisions [they] actually care about." [3]. Still, other varieties are 'soft' in that they preserve freedom of choice and are thus fundamentally libertarian. This freedom of choice is paired with influencing choices and, as a result, the effect is a guiding hand. In continuation of this we understand a nudge as an implicit advice to a specific act.

Nudging uses insights into behavioural economics to simplify the choices and their consequences [19, 13, 40, 53, 49]. As John et al. [19] argue, the bounded human rationality influences the interaction between the cognitive structure, such as heuristics, and the perception of the choice environment. Choice architecture frames choices because it is the background against which decisions are made [49]. The environment influences perception,

interpretation and action in the choice situation. Nudging aims to overcome the rational inadequacies of an individual by designing an environment that intervenes in the individual's decision making. This is why nudging has the potential to change behavior without an active reflection by the citizens.

In the following we define provocative communication with critical design as the starting point and include preliminary results from an experiment using this style of communication.

2.2 Provocative Communication

Bardzell and Bardzell [1] have pointed out that critical design is a research through design methodology, which is grounded in ethical considerations and can reveal potentially hidden agendas and values, and explores alternative design values [1]. Bardzell and Bardzell argue that Dunne and Raby give only little methodological guidance to working with critical design and propose that further examples is needed to understand the methodological aspects better. Following this suggestion we, in this paper, discuss its application in an environmental context, that of waste minimisation. So, we propose a variant of critical design called provocative communication framed by a digital media context, namely social media.

Critical design has been coined by Anthony Dunne and Fiona Raby [7] and the term covers the practice of designers with a professional ethical stance. In this sense critical design is a form of research, which includes the design of a product; a broad definition, that invites its users, in this case citizens, to critically reflect upon how their everyday lives are influenced by hidden assumptions, values, ideologies, and behavioral norms inscribed in the designed world [7, 6].

As noted by Bardzell and Bardzell [1] critical design aims at developing a critical reflection within the citizens towards 'reification' of society. Reification "refers to the way that things are produced by society, including the way that it is organized, appear as entirely natural and beyond question." [15].

Critical design aims at raising reflection and a certain critical sensibility within the citizens, which Dunne and Raby define as: "The critical sensibility, at its most basic, is simply about not taking things for granted, to question and look beneath the surface. This is not new and is common in other fields; what is new is trying to use design as a tool for doing this." [8].

Inspired by critical design, provocative communication presents information in a thought-provoking way to enable a critical reflection by the parties involved in the communication. To our knowledge the term 'provocative communication' as a critical design type has yet to be defined and so we will give a preliminary definition still open to further refinement.

To Dunne and Raby [7] the nature of design is ideological being either affirmative or critical. Affirmative design reinforces the existing situation, not questioning the underlying ideologies. Contrary to this, critical design objects the current situation by using design as a critique to produce alternative values, ultimately transforming society's ideologies. Provocative communication embraces the latter in order to appeal to System 2 of the communicating parties. This communicative style refrains from pointing to ends or solutions; it only questions existing issues, leaving the solutions to the parties involved. In this way, provocative communication is comparable to the think strategy proposed by John et al. [18, 19], but differs by including both citizens and authority representatives, and initiating the communication by framing the issue at hand in a thought-provoking manner. Thus, it should initiate a deliberative debate, resulting in behavioural change.

As part of Nulskrald's phase 3 a six week experiment (May to June, 2015) by the first author of this paper used provocative communication. Five families with children participated in the experiment that focused on reducing food waste in the household and the data was collected using diaries, interviews and workshops. The communicative style was tested by sending a weekly SMS for a week, containing food waste facts . These made the families reflect on their own practices to a greater extent which also motivated them to change and improve their practices. This is a preliminary result from food waste experiment. An extended use of provocative communication on Facebook and elsewhere is currently being planned.

Whilst provocative communication is a design approach to communication to create mutual critical reflection on the issues in question, nudging intends to steer behaviour without deliberation.

As we will show in our discussion, provocative communication needs facilitation by the 'I', in our case AVV. Because of the provocative nature of the communication, 'the other person' is, at least potentially, forced to reflect upon what has been said and done. A reflection, which opens new possible venues for action and reaction that makes it possible for 'the other person' to contribute to and be acknowledged in the participatory setting of social media or online discussions. At the same time, AVV would have to meet the citizens on their terms, provide them with relevant feedback and input, to be able to recognise and facilitate a possible movement or community.

DISCUSSION

In a larger perspective, nudging is a top-down approach used by governments and municipalities [11]. This means that for choice architecture to work and be accepted by the citizens, the choice-architects must be credible in the eyes of the public [29,5,39]. At the same time, participatory culture, as seen on the Internet, demands relevant input for the participants, as well as recognition and feedback on the participant's input [16]. Default strategies frequently used by municipalities and governments include policy instruments such as regulations, laws and taxes [9]. Such strategies are in Sunstein's [47] words 'hard' and designed to influence people's choices of ends. Thus, they involve a strong focus on paternalism, not leaving room for individual deliberation which excludes the ideas of empowerment and freedom. The notion that nudges aim at influencing and exploiting the automatic system – that is, System 1 – especially substantiates this issue.

Nudging works on an individual level [19]. Additionally, a nudge always influences automatic behaviour, but rarely influences deliberate choices [13]. In like manner, Selinger and Whyte [40] state, that "[a] nudge does not try to inform the automatic system, but work with the influence biases inevitably have." In this way, nudging is natural design that affords a specific behaviour without involving reflective thought [30]. Goodwin [11]) argues, that part of the motive to apply nudging practices to British policy is to empower citizens and, moreover, to promote freedom and fairness. Furthermore, Goodwin [11] questions the libertarian aspect of nudges by pointing to the issue that the optimum working conditions for a nudge is when people are unaware that their choices are influenced. Some nudges, however, involves System 2 [49, 47], but they are still paternalistic.

As we have pointed out above, Løgstrup as well as critical design demands another approach to the behavioural change of citizens. An approach, which means a larger involvement of AVV or other

organisations like them, as well as an on-going facilitation of citizens in acknowledging their input. All of this constitutes a learning environment, which enables and forces reflection and hereby giving rise to change.

In the following we discuss the view that nudging embraces empowerment, freedom, and fairness in relation to the ethical demand including reflections on subjective well-being which stands in opposition to government's determination of what makes people's lives better.

3.1 Empowerment, Freedom, and Fairness

The contestation that empowerment is a part of a nudge gives cause for ethical concerns since a nudge is designed to affect System 1, thus, not involving System 2. As Waddock [52] shows, empowerment can be defined differently depending on the context in which the term is used. In this paper we employ two of Waddock's definitions. The first is when an authority gives power to individuals or a group in a subordinate position. The second definition of empowerment is a specific "psychological state of mind for individuals or groups, which allows them to feel a degree of control over their own goals and accomplishments." [52]. The feeling of empowerment motivates people to accomplish goals and is underpinned by the feeling of "self-determination, self-efficacy, and capability to bring about impacts or changes" [52]. Both definitions imply the concept of autonomy since empowering individuals and groups enables them to act on their own corresponding to their subjective goals. In the context of libertarian paternalism, the Millian definition of autonomy is closely related to the libertarian aspect given that people are autonomous when they choose to frame their plan of life for themselves and to do what they want [28]. Thus, empowerment hinges on autonomy which in turn hinges on liberty or freedom.

AVV's ambition to start a movement should then include empowerment of citizens by facilitating the development of a self-sustaining community. However, nudging is not capable of motivating citizens to create a community. As a policy tool to steer behaviour, nudging can, thus, be used to guide the movement in a certain direction in accordance with AVV's vision to *Motivate to minimise waste*. Empowering and steering at the same time subverts citizen's feeling of autonomy, especially if nudging is used as motivator; a tool to empower citizens.

The empowerment can be viewed as an illusion since the benefits of a nudge not necessarily concern the citizen. Actually, some nudges are designed with a utilitarian end goal only taking society into consideration. In doing so, there is a strong focus on paternalism making the nudge coercive, that is, to use force to get an individual to act in a way that serves another person's will for the other's purpose [37]. A coercive nudge is labelled libertarian, but the restriction of options in the choice situation makes it paternalistic. Although AVV could understand a nudge as libertarian, the citizen could feel his or her freedom as restricted. Hence, the citizen can only choose not to act according to AVV's guiding hand if System 2 is utilised in the situation; something that the very nature of the design of the nudge decreases the probability of happening. In this way, the individual would have to be able to figure out the AVV's motive and end goal with the nudge, requiring complete transparency of the design and rational deliberation (System 2). But how does AVV favour transparency as one of the three missions mentioned above is *Transparent business*?

Applying provocative communication ensures transparency since it implies a mutual and respectful deliberation. In such a communicative act all participating parties are equal in that they have a responsibility to provide their own knowledge and opinions while not oppressing the others' life manifestations. Moreover, a strategic application of provocative communication focusing on e.g. food waste facts could enable a basis for a nascent community.

Contrary to nudging, provocative communication needs constant renewal on e.g. the social media to keep people engaged in the deliberative process. Otherwise the communication and with it the community would simply cease to exist. Thus, AVV must identify itself with the citizens to keep the deliberation of possible solution to minimise food waste going. The relationship between AVV and the citizen – the 'I' and 'the other person', respectively – must be based on mutual respect and trust. Thaler and Sunstein [49] argue that if this respect is not manifested then citizens are treated as means to a goal. When using nudges or provocative communication, AVV must in addition be able to defend these interventions in public. Otherwise, "it treats its citizens as tools for its own manipulation." [49]. Hence, transparency denotes openness and intelligibility. Moreover, the constant renewal of the communication enables organisational learning as the relationship between AVV and the citizens is dynamic and provides new perspectives on the issue at hand through the communication. Provocative communication entails a much closer relationship to the citizens because of the openness to AVV as an organisation, making AVV reflect on itself.

From an ethical point of view, provocative communication maintains the ethical demand by honouring empowerment, authority, and fairness. This is not the case for nudging. Following Hausman and Welch's [14] argument, a nudge is coercive as it undermines the deliberation and liberty of the citizen by aiming at the automatic behaviour. Consequently, the citizen is not in control. Moreover, Hausman and Welch [14] argue that it is by focusing only on the contents of the choice set that a nudge can be counted as 'libertarian'. Seen from a libertarian paternalistic perspective, nudges re-arrange choice options without forbidding any. In this sense, no nudges are paternalistic, as Hausman and Welch [14] point out. A nudge is only paternalistic if it limits the options available [14]. Paradoxically, the transparency and monitorability of the designer's motives becomes more intelligible if it is based on hard paternalism [38]. Rebonato's [38] discussion highlights this issue. Thaler and Sunstein [49] points out that "[t]here is no such thing as 'neutral' design […] Small and insignificant details can have major impacts on people's behaviour. A good rule of thumb is to assume that 'everything matters.'" [49]. In line with Rebonato [38], it is questionable if a nudge truly enables a chooser to pursue his or her personal preferences in the choice situation.

The fairness of nudges is dubious. As a policy instrument, libertarian paternalist nudges presents themselves as interventions that affect the means by which people try to reach their ends. Nonetheless, they are designed with a specific end goal in sight that reduces society's major ills [49]. This conflicts with the ethical demand [24] because the citizen ('the other person') is not empathically acknowledged by the authority (the 'I') and is hindered in achieving his or her full potential as a human being. Without knowing the personal end goals of every citizen, a nudge can only be designed on a general understanding of citizen's needs and end goals. As a result, the 'I' is unable to take 'the other person's' subjective well-being into account when designing nudges. A nudge is only truly empowering, libertarian, and fair from the nudger's perspective, holding a narrow conception of these factors.

CONCLUSION

When changing citizen's behaviour to minimise waste and build a community, AVV is confronted with ethical challenges of which role to play and the organisational implications. These derive from the behavioural design approaches employed by AVV. The ethical issues are manifold and in this paper only some of them have been discussed.

The ethical challenges posed by nudging, includes the asymmetrical distribution of power, since AVV is the one articulating in what direction to steer the citizen's behaviour. For the main part, a nudge is only libertarian when seen from the nudger's perspective, while a nudge is mostly paternalistic when seen from the citizen's perspective. Thus, with a subjective, narrow conception of empowerment, autonomy, and freedom as staple elements of a nudge, the 'I' cause ethical implications for the relationship with 'the other person'. The design of a nudge runs counter to the ethical demand by exploiting people's cognitive limitations. A designed nudge needs no further refinement and is left to itself, maintaining the behaviour caused by the nudge. On the whole, nudging oppresses the citizen's – or 'the other person's' – life manifestations. To comply with the ethical demand as it is posed by 'the other person' towards the 'I' another design approach is needed.

Thus, provocative communication is another possibility, since it seems to respect and acknowledge 'the other person's' life manifestations. This is due to the transparency, which implies a mutual and respectful deliberation based on the idea that people are equal in that they have a responsibility to provide their own knowledge and opinions while not oppressing the others' life manifestations. Hereby provocative communication maintains the ethical demand by honouring empowerment, authority, and fairness.

Concerning organisational learning and development, nudging does not require any, once the nudge is in place. In fact, nudges will typically manifest the knowledge already inherent in the organisation. Provocative communication on the other hand, requires learning, development, as well as facilitation of an on-going discussion, acknowledgement, and change in behaviour for both organisation and citizen. Thus, nudging is the easy way out, while provocative communication requires constant work and evolution.

ACKNOWLEDGMENTS

Our thanks to AVV and the five families for participating in the food waste experiment.

REFERENCES

[1] Bardzell, J. and Bardzell, S. 2013. What is "Critical" About Critical Design? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3297–3306). New York, NY, USA: ACM. DOI= http://doi.org/10.1145/2470654.2466451

[2] Chaiken, S. and Trope, Y. (Eds.). 1999. *Dual-process theories in social psychology*. Guilford Press.

[3] Conly, S. 2014. *Against Autonomy: Justifying Coersive Paternalism*. Cambridge University Press.

[4] Danish Environmental Protection Agency. 2012. Affaldsbekendtgørelsen - Bekendtgørelse om affald. (April 2012). Retrieved June 30, 2015 from: https://www.retsinformation.dk/Forms/R0710.aspx?id=144826

[5] Druckman, J. N. 2001. Using Credible Advice to Overcome Framing Effects. *Journal of Law, Economics, and Organization*, 17, 1, 62–82. DOI= http://doi.org/10.1093/jleo/17.1.62

[6] Dunne, A. 2006. *Hertzian Tales: Electronic Products, Aesthetic Experience, and Critical Design*. MIT Press.

[7] Dunne, A. and Raby, F. 2001. *Design noir: The secret life of electronic objects*. Springer Science & Business Media.

[8] Dunne, A. and Raby, F. 2009. Interpretation, collaboration, and critique: Interview with Dunne and Raby. Retrieved July 1, 2015 from: http://www.dunneandraby.co.uk/content/bydandr/465/0

[9] Firestone, J. 2002. Agency governance and enforcement: the influence of mission on environmental decisionmaking. *Journal of Policy Analysis and Management*, 21, 3 (June, 2002), 409–426. DOI= http://doi.org/10.1002/pam.10052

[10] Gilovich, T. and Griffin, D. 2002. "Introduction – Heuristics and Biases: Then and Now". In Gilovich, T., Griffin, D., and Kahneman, D. *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

[11] Goodwin, T. 2012. Why We Should Reject "Nudge." *Politics*, 32, 2 (May, 2012), 85–92. DOI= http://doi.org/10.1111/j.1467-9256.2012.01430.x

[12] Greimas, A. J., 1966/1983. *Structural Semantics: An Attempt at a Method*. McDowell, D., Schleifer, R., and Velie, A. Lincoln (trans.). Nebraska: University of Nebraska Press.

[13] Hansen, P. G. and Jespersen, A. M. 2013. Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. *European Journal of Risk Regulation: EJRR*, 4, 1 (April, 2013), 3–28.

[14] Hausman, D. M. and Welch, B. 2010. Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, 18, 1 (November, 2009), 123–136. DOI= http://doi.org/10.1111/j.1467-9760.2009.00351.x

[15] How, A. 2003. *Critical Theory*. Palgrave Macmillan.

[16] Jenkins, H., Ford, S., & Green, J. 2013. *Spreadable media: Creating value and meaning in a networked culture*. NYU Press.

[17] Jensen, T. 2013. Designing for relationship: Fan fiction sites on the Internet. *Teoretisk og Anvendt Etik (Theoretical and Applied Ethics)*, 5, 1, 241-255.

[18] John, P., Smith, G., and Stoker, G. 2009. Nudge Nudge, Think Think: Two Strategies for Changing Civic Behaviour. *The Political Quarterly*, 80, 3 (August, 2009), 361–370. DOI= http://doi.org/10.1111/j.1467-923X.2009.02001.x

[19] John, P., Cotterill, S., and Richardson, L. 2011. *Nudge, Nudge, Think, Think*. Huntingdon, GBR: Bloomsbury Academic. Retrieved July 1, 2015 from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10511470

[20] Jones, R., Pykett, J., and Whitehead, M. 2011. The Geographies of Soft Paternalism in the UK: The Rise of the Avuncular State and Changing Behaviour after Neoliberalism. *Geography Compass*, 5, 1 (January, 2011),

50–62. DOI= http://doi.org/10.1111/j.1749-8198.2010.00403.x

[21] Keshavarz

[22] Liquid Agency. 2012. The Strategic Pyramid. Retrieved July 2, 2015 from: http://www.liquidagency.com/blog/the-strategic-pyramid/#.VL44i0eG98E

[23] Løgstrup, K. 1988. *Udfordringer*. Hadsten: Mimer.

[24] Løgstrup, K.E. 1997. *The Ethical Demand*. Notre Dame: University of Notre Dame Press.

[25] Løgstrup, K. E. 2007. *Beyond the Ethical Demand*. Notre Dame: University of Notre Dame Press.

[26] Løgstrup, K. E. 2014. *Etiske begreber og problemer*. Aarhus: Forlaget Klim.

[27] McDougall, F. R., White, P. R., and Franke, M. 2008. *Integrated Solid Waste Management: A Life Cycle Inventory*. Chichester, GBR: Wiley. Retrieved July 1, 2015 from http://site.ebrary.com/lib/aalborguniv/reader.action?docID=10240521

[28] Mill, J. S. 2001. *On Liberty*. London, GBR: ElecBook. Retrieved from http://www.ebrary.com

[29] Moseley, A. and Stoker, G. 2013. Nudging citizens? Prospects and pitfalls confronting a new heuristic. *Resources, Conservation and Recycling*, 79 (May, 2013), 4–10. DOI= http://doi.org/10.1016/j.resconrec.2013.04.008

[30] Norman, D. 1988. *The design of everyday things*. New York: Basic Books.

[31] OECD. 2004. Working Group on Waste Prevention and Recycling. (September 2004). Retrieved June 30, 2015 from: http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/epoc/wgwpr/se(2004)1/final

[32] OECD. (n.d.). Waste prevention and minimisation. Retrieved June 30, 2015, from https://www.oecd.org/env/waste/prevention-minimisation.htm

[33] Pahuus, M. 2000. *Holdning og spontaneitet: Pædagogik, Menneskesyn og værdier*. Århus: KvaN.

[34] Piaget, J. 1950/2005. *The psychology of intelligence*. London: Routledge.

[35] Piaget, J. and Inhelder, B. 1969/2000. *The psychology of the child*. Basic Books.

[36] Poulsen, S. B. and Strand, A. 2014. A creative designerly touch. *Academic Quarter*. Spring edition.

[37] Price, T. 2008. Coercion. In R. Hamowy (Ed.), *The encyclopedia of libertarianism*. (pp. 76-77). Thousand Oaks, CA: Sage Publications, Inc. DOI= http://dx.doi.org/10.4135/9781412965811.n50

[38] Rebonato, R. 2014. A Critical Assessment of Libertarian Paternalism. *Journal of Consumer Policy*, 37, 3 (August, 2014), 357–396. DOI= http://doi.org/10.1007/s10603-014-9265-1

[39] Riker, W. H. 1995. The Political Psychology of Rational Choice Theory. *Political Psychology*, 16, 1 (March, 1995), 23–44. DOI= http://doi.org/10.2307/3791448

[40] Selinger, E. and Whyte, K. 2011. Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture. *Sociology Compass*, 5, 10 (October, 2011), 923–935. DOI= http://doi.org/10.1111/j.1751-9020.2011.00413.x

[41] Simon, H. A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics* 69, 1 (February, 1955), 99–118. DOI= http://doi.org/10.2307/1884852

[42] Simon, H. A. 1979. Rational decision making in business organizations. *The American economic review*, 69, 4 (September, 1979), 493-513.

[43] Stanovich, K. E. and West, R. F. 2000. Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 5 (October, 2000), 645–665.

[44] Starke, L. (ed.). 2012. *State of the World 2012: Moving Toward Sustainable Prosperity*. Island Press, Washington, DC, USA. Retrieved July 2, 2015 from http://www.ebrary.com.

[45] Strand, A. M. C. 2014a. Material Storytelling. Resituating Language and Matter in Organizational Storytelling. In Jørgensen, K. M. & Largarcha-Martinez, C. (Eds.). *Critical Narrative Inquiry – Storytelling, Sustainability and Power in Organizations*. New York: Nova Science Publishers.

[46] Strand, A. M. C. 2014b. The Story of Grandma's Dress (code): Practices diffracted through the Apparatus of Material Storytelling. In Boje, D. M. and Henderson, T. (Eds.). *Being Quantum. Storytelling and Ontology in the Age of Antenarratives*. Newcastle: Cambridge Scholar Publishing.

[47] Sunstein, C. R. 2014. The ethics of nudging. (November, 2014). Available at SSRN. DOI= http://dx.doi.org/10.2139/ssrn.2526341

[48] Thaler, R. H. and Sunstein, C. R. 2003. Libertarian Paternalism. *The American Economic Review*, 93, 2 (May, 2003), 175–179.

[49] Thaler, R. and Sunstein, Cass R. 2009. *Nudge: Improving decisions about health, wealth, and happiness* (Rev. and expanded ed.). London: Penguin Books.

[50] Tversky, A. and Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 4157 (September, 1974), 1124–1131.

[51] United Nations Environment Programme (UNEP). (n.d.). Global Partnership on Waste Management (GPWM). Retrieved June 15, 2015, from http://www.unep.org/gpwm/Background/tabid/56401/Default.aspx

[52] Waddock, S. 2008. Empowerment. In R. Kolb (Ed.), *Encyclopedia of business ethics and society*. (pp. 715-716). Thousand Oaks, CA: SAGE Publications, Inc. DOI= http://dx.doi.org.zorac.aub.aau.dk/10.4135/9781412956260.n280

[53] Wilkinson, T. M. 2013. Nudging and Manipulation. *Political Studies*, 61, 2 (June, 2013), 341–355. DOI= http://doi.org/10.1111/j.1467-9248.2012.00974.x

[54] Winkel, T. D., Bolvig, S., and Rosenstand, C. A. F. 2015. The Challenge of a Sustainability Change. Nordic Design Research (NORDES). 6.

Ethical Competence and Social Responsibility in Scientific Research using ICT Tools

Ryoko Asai
Uppsala University
Box 337, SE-751 05
Uppsala, Sweden
ryoko.asai@it.uu.se

Iordanis Kavathatzopoulos
Uppsala University
Box 337, SE-751 05
Uppsala, Sweden
iordanis@it.uu.se

ABSTRACT

This study explores how to improve and support researchers' ethical competence in scientific research and how to conduct research ethically, especially in research activities using Information and Communication Technology (ICT). Refining research ethics relating to ICT is unavoidable in the highly technological society of today, for example big data is used in different scientific research activities, and systems which support our daily lives are constructed based on the existing systems. In other words, technology reproduces technology itself. And almost all research activities need to use ICT through the whole research process. Moreover, researchers are required to be able to participate and react sensibly in ethical dialogues with society and citizens. Seen in that light, this study could be applicable not only to computer science and technology but also to a broad spectrum of research areas as the constructive notions of ethics, liberty and responsibility in research activity.

Categories and Subject Descriptors

K.4 [**Computers and Society**]: Ethics; K.3.m [**Computers and Education**]: Miscellaneous—*Computer literacy*

General Terms

Theory

Keywords

Autonomy, decision making, ethical competence, ethical guidelines, research ethics

1. INTRODUCTION

These days, we have seen many academic scandals through media, sometimes in person. Many of those scandals are related to ethical issues, for examples plagiarism, falsification, intrusion of privacy etc. Development of Information and Communication Technology (ICT) contributed to scientific research greatly and it reproduce more advanced technology. On the other hand, new technology brought new conflicts and issues to research activities. Today, researchers are required to be able to participate and reply sensibly in ethical dialogues with society and citizens. This study explores how scholars could be ethical from the philosophical and sociological perspectives. And we also examine the usefulness of ethical codes and guidelines and the potential risk of ethical codes and guidelines.

2. SOCIAL ROLES OF SCHOLARS

At the present day, almost all research activities cannot avoid using ICT as a tool for writing papers, checking research sources, contacting others, analyzing data, coding, undertaking statistical analyses, setting up and planning research topics and so on. ICT is applied to all kinds of research activities and areas. It may be no exaggeration to say that scientific research activities are supported greatly by ICT. Using ICT skilfully has become a vital part of research activities.

Among a large number of research areas, scholars in computer science often handle huge amounts of data, especially personal data under the label of scientific/academic research. And their research efforts to research contribute to improving existing technologies, creating innovations and pushing new technology to adapt to our daily lives. In other words, they take an important social role to design and shape society for the future through their research activities. Because of a researcher's important social role, researches and researchers are required to be ethical and also to follow the rules which are established on each research area.

3. CLASH BETWEEN NEW TECHNOLOGIES AND THE EXISTING RESEARCH ETHICS

Needless to say, scientific research activities have been conducted under the strict rules and research ethics guidelines, which are strongly established in every research area. Generally, scholars need to understand the research guidelines covered broad research activities and also the research ethics codes which are sometimes described vaguely. In some cases, the general guidelines are difficult to apply to research activities or don't work properly. The research areas where directly influence on human life like biology, the more stringent rules are imposed upon research activities. In biotechnology area, it is well known that the research on human cloning has posed considerable ethical and moral issues since late 90's.

Even if technological feasibility is higher and an astonishing result is expected, it is believed that there are socially unacceptable research activities from the perspective of ethics. On the other hand, with the development of technologies, scientific researches are getting more interdisciplinary. Project-based research employs scholars from a varied of research fields. Moreover, industries provide huge amount of research funding to research organizations, and also collaboration between industries and academia is very common and active today. In those cases, researchers who have different research guideline and ethics work together in the same research project. How do they share the wide-ranging research guidelines and ethical conduct? Which rules should they follow? How do researchers discipline themselves ethically in conducting their research?

4. ETHICAL ISSUES OF USING TECHNOLOGY IN SCIENCE

In leading a research project in biology and medicine, but also in other disciplines, we use the latest computer tools to handle and treat our data. Usually the amount of data gathered is enormous and in order to be able to grasp them and make them meaningful a bioinformatician may be engaged to take care of the data, for example by the creation of an algorithm to systematize the data. However, this operation transforms the richness of data to a few simple categories. The problem is that if the results are presented in this simplified way there may be misinterpretations that will misguide future research. On the other hand it is clear that the scientists can never get their research published unless they simplify the data.

Those ethical issues could happen not only in bioinformatics and natural science but also in social science and humanity. In 2014, some "socially unaccepted" or "unethical" research results were published in scientific journals, even passing the peer-review process. Especially a research paper on psychological experiments using Facebook provoked a big controversy not only in public but also in academia [1]. The researchers belong to Facebook as a researcher or got research support from Facebook [1]. Because of that, they could use Facebook interface and its big data for their psychological experiments. The results of their experiments seem to be useful and interesting in developing online contents and improving their usability.

However, the public and many other researchers criticized that the authors, the research and the journal on which published it were unethical. Although their research may be seen as legal [2], still it is criticized as unethical and it is not accepted by society. Directly after big criticism over Facebook's phycological experiment, the authors explained and emphasized that their experiment on Facebook did never violate laws and also agreements with users. when seeing the gap between the public and the authors, we recognize that 'following rules and laws' and 'being ethical' are different.

Furthermore, our history and archives also ask researchers how to research ethically. These days, not only big date and social networks but also information policies have developed dramatically. With those technologies' development, governments in many countries have organized their laws and policies. Today, the date and results of Nazi experiments evokes an argument if it is ethical to use Nati medical experiments and how to use the date if it is ethical [3]. I was obvious that those results and data of Nati experiments were conducted under inhumane and unethical conditions and many people lost their lives because of those experiments during World War II. Can we justify to use the data and knowledge from Nati experiments? If we follow the utilitarian direction which focuses on consequence, we could justify to use those results and data to make society better and enhance medical technology and the quality of life. However, if we focus on the process how to get those result and data, the different decision might come out, even if those data and knowledge contribute to the future. The most important issue is how each scholars justify their own research ethically and how they take a responsibility when they face the ethical conundrum. How can they acquire a strong conviction that their research work is ethical? Where is the line between ethical and unethical research?

5. IMPROVING ABILITY TO MAKE A DECISION ETHICALLY

Scholars assume great responsibility for society and for the future through their research activities. A huge number of scholars use ICT as a research tool and work on developing these tools. Some of them have great opportunities to handle and analyze or users' private information in research activities. On the other hand, when they face ethical issues in their research they are confounded by equivocal rules, ambiguity of ethical codes and a lack of ethical competence. Basically business ethics is not applied to academic research, although we can observe some commonalities between them. However, nowadays, many researchers work in industries and it is not hard to imagine they get more used to ethical conduct in business area. And it is also highly possible that new technology might create new ethical conflicts in research.

5.1 Two of liberties in working on scientific research

Scholars are always required to conduct research ethically and to contribute to the basic needs of society. Moreover, research activities aim to be rational, and researchers are required to be independent, to keep their positions neutral, and to take a balance between public benefits and their own benefit. Simply saying, scholars take a great responsibility for society and the future. However, if society imposes them

[1]"Furor Erupts Over Facebook's Experiment on Users" (posted on June 30, 2014) by The Wall Street Journal, http://www.wsj.com/articles/furor-erupts-over-facebook-experiment-on-users-1404085840, and also see "Facebook emotion study breached ethical guidelines, researchers say" (posted on June 30, 2014) in the guardian, http://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say

[2]In September 2014, two law professors in University of Maryland alleged the social network experiments by Facebook and OKcupid violate state statute and announced their opinion in public. See more: http://james.grimmelmann.net/files/legal/facebook/MDAG.pdf

[3]"Is it ethical to use data from Nazi medical experiments?" by The Conversation, http://theconversation.com/is-it-ethical-to-use-data-from-nazi-medical-experiments-39928

strong restrictions, their creativity and uniqueness would be denied and society might lose an opportunity to enhance technology and the quality of life in the future. Scholars need to have liberty to work on their own research activities based on their interests, values and intelligence as well as to take a responsibility for society. In other words, scholars are required to take a balance between *positive liberty* and *negative liberty* [2].

when scholars focus on *positive liberty*, they can work on research based on their own motivation, interest and values. They could behave by their own will and decision. However, if *positive liberty* allows all of them to work based on their own will and decision, many and varied interests and values might clash and violate others' will and decision. They have a liberty to work on their research though, rules and ethical codes are needed. When they place themselves in *negative liberty*, they could work on research without violation and restriction. Even if there are social pressure and restrictions on research, they have a liberty from those pressure and restrictions and they work on their research without interfere.

5.2 Ethics of responsibilty

Creative and advanced researches are driven by scholars' strong motivation and will for research. But strong motivation and will have a risk to lead ethical issues, and they allow others to interfere their research by rules and restrictions. At the present day, it is impossible to suppose that research is conducted without any interference. Basically researchers are constrained by many social factors such as rules, values, and ethical codes. In that sense, researchers need to have *negative liberty* to work on their creative and novel research. Both sides of liberties are needed for research activities. And researchers are required to control two of liberties.

Scholars need to make a decision personally in some cases through the long research activity. Or sometimes the research might be conducted by one researcher. Under those situations, their choice might be seen as a personal choice. However, as long as their personal choice is ensured by liberty and they can control two sides of liberties, which is essentially made for achieving good or common good in society [3] [4]. Society assumes ultimate responsibility even for personal choice [5]. It is very difficult for self-indulgent or asocial decision to be accepted by others and society without social validity and social approval. Therefore, even personal decision-making tends to contemplate for good intentionally or unintentionally.

Weber described that there are two different ethical maxims: one is *ethics of conviction* and the other is *ethics of responsibility* [5]. A person standing upon *ethics of conviction* would feel responsibility for his/her pure conviction. However, as long as the action derives from pure conviction, s/he has little interest in accepting the blame for an undesirable consequence. In an ethics of conviction, there is a risk of restricting or violating personal liberty. It corresponds to the situation where a person pursues positive liberty extremely. On the other hands, a person focusing on *ethics of responsibility* assumes responsibility for foreseeable consequences of actions. What matters to scholars who take a part in designing society is responsibility for the future [5]. Therefore, scholars are required to have ability to foresee the future and make a decision supported by *ethics of responsibility* .

6. ETHICAL CODES AND GUIDELINES

Ethical codes may be helpful in preventing and handling ethical problems as well as taking a balance between *positive liberty* and *negative liberty*. However they may create additional problems to the organization and to the persons using them.

The ideal ethical code should provide cognitive support during the effort to think autonomously; help the person and the group/organization to think and keep the dialogue on a philosophical level. It can be also a toll for training of autonomy skills during the formulation, interpretation and revision processes of the code. It can support democratic communication and dialog, establish autonomous structures and processes in the organization and the group, it can be used as a tool for guidance, to support anticipation and planning, and to solve conflicts or remove the causes for conflicts before they emerge. Ethical codes may help persons and groups/organizations to turn focus on own responsibility by expressing contradictions and inconsistencies in its rules, and they can promote confidence, on both a personal and a group level, by offering a way to handle moral issues.

There are however risks involved in the use of ethical codes. Any rule or guideline not being included in the code may be interpreted as morally allowed. Or something stated there but not fit for a certain situation may be seen as morally compulsory and being enforced nevertheless. An ethical code may become a weapon in conflicts, a proxy for any kind of conflict. It can consolidate current moral values, strengthen and shield moral correctness, hinder change and adaptation to new conditions. Ethical codes and guidelines may support the creation of moral facades, and facilitate career making; promote the establishment of moral hierarchies, structures and procedures; strengthen heteronomy and hinder autonomy at personal and group levels; and shift responsibility from persons and groups to the rules themselves.

7. ACKNOWLEDGMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006.

8. REFERENCES

[1] A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[2] I. Berlin. *Four essays on liberty*, volume 969. Oxford University Press New York, 1970.

[3] A. D. Bloom. *The republic of Plato*. Basic Books, 1991.

[4] H. Nagai. *What is ethics?* Chikuma shobo, Tokyo, 2011.

[5] M. Weber, D. S. Owen, and T. B. Strong. *The vocation lectures*. Hackett Publishing, 2004.

What is required of requirements? A first stage process towards developing guidelines for responsible research and innovation

Sara H. Wilford
Centre for Computing and Social Responsibility
De Montfort University
G.H 5.77 Gateway House
Leicester
(+44) 0116 2506294
sara@dmu.ac.uk

ABSTRACT

Responsible research and innovation (RRI) considers the impact of development on stakeholders and provides a direction for the future of science and technology. Therefore, in the practical world of the lab, what is needed is a set of guidelines to assist in the application of those RRI principles. However, to ensure that any guidelines are usable and acceptable, it is important to engage with those who would actually be expected to implement them.

Stakeholders are often asked to evaluate a set of guidelines or recommendations without having any say in how they are constructed, what they should look like or what they should contain. The process of stakeholder engagement in the development of a set of 'requirements' therefore provides insight from which a set of guidelines can be developed. In this way, acceptance is fostered through stakeholder involvment in the process, which has been built from the core principles of RRI.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Ethics

General Terms

Management, Performance, Design, Human Factors, Standardization, Theory.

Keywords

Guidelines, Requirements, RRI, Participation, Stakeholders.

1 INTRODUCTION

Taking personal responsibility for our actions and the impact of those actions is something we are taught from an early age. We are expected to be honest, admit our mistakes and rectify and/or apologise where we do harm.

Responsibility in the context of research and innovation and as a key element of RRI blurs the boundaries between the personal and the institutional. Being responsible 'to' can involve a chain of command or similar whereby the lines of responsibility for completion of a task or some other obligation is directed towards an individual or an organization, often through specific channels of communication. This form of responsibility can lead to 'passing the buck' and may allow individuals to avoid taking personal responsibility. Being responsible 'for' something however remains with the personal and includes taking responsibility for the outcomes of one's actions, and a concern about those who are likely to be affected both within and beyond an organization.

RRI re-engages the individual with personal responsibility at the same time as re-inforcing institutional responsibility. This means that RRI creates a step-change in the way that those who are engaged in research and innovation should consider the impact of what they do. To encourage RRI take-up amongst researchers and innovators across all sectors therefore, guidelines and recommendations are needed to provide a starting point for its adoption. However, guidelines for the governance of RRI need to be broad enough to encompass all stakeholders and yet flexible and specific enough to enable stakeholders to frame their own particular contextual understanding of RRI. Indeed, if there is to be any hope of success in normalizing the key principles of RRI into the working practices of researchers and innovators, it will not be through rigid and inflexible approaches.

This paper addresses part of the process in developing a set of guidelines and recommendations for the governance of RRI. Creating a set of guidelines are often key requirements of research projects and can be aimed at a wide audience ranging from researchers and civil society organisations (Stahl and Wakunuma, 2015) to project co-ordinators (Fedor et al, 2006).

Of the key pillars of RRI, participation and stakeholder engagement (Pelle and Reber 2013) are considered to be particularly important. Therefore, in order to create a set of guidelines it seems logical to involve stakeholders at each step of the process to have a clear idea about what the guidelines should look like and the nature of the content to be included.

When creating guidelines, this engagement generally occurs at the point after the first draft has been constructed to enable stakeholders to evaluate and revise them before finalizing. Decisions about what guidelines should contain and how they should be presented are generally taken in the first instance by their creators and presented to stakeholders as a *fait accompli*.

The research aimed at developing a set of guidelines for RRI in practice and across a broad spectrum of needs and concerns. Althought chiefly aimed at researchers and innovators, the guidelines may also act as a guide to other stakeholders to better understand the principles under which they should be working if they are to comply with RRI. This may be particularly relevant to those seeking funding from national (public) funding institutions such as the European Commission and other bodies such as the EPSRC which has recently adopted the AREA framework; *Anticipate, Reflect, Engage, Act* (Stilgoe et al 2013) to promote RRI within its mission.

This paper presents the process of stakeholder engagement that should occur before the guidelines are created and addresses an important gap in the process i.e the requirements for guidelines. By using this approach the requirements and subsequent guidelines are likely to have greater validity and acceptability to those expected to use them.

The paper firstly considers RRI and its importance and relevance to future developments and then considers guidelines in context and how norms, governance and reflexivity are critical factors in establishing a set of guidelines that will be useful and relevant. The rationale behind the approach to the requirements for the guidelines is discussed and then the methodology and process is detailed. Finally, the paper concludes by indicating how user developed guidelines for guidelines can inform the creation of the guidelines themselves and that the process can be utilized in other projects where the development of guidelines are a required element.

2 GUIDELINES IN CONTEXT

2.1 Responsible research and innovation

In general terms, responsible research and innovation (RRI) describes how research and innovation in all fields of endeavour, can be beneficial to stakeholders by considering possible impacts from the outset. The idea that all fields including management, science, sociology, ethics and engineering could each strive towards the same ultimate goal under an umbrella of RRI has grown in recent years, (Stahl et al, 2013; Owen et al, 2012; Sutcliffe 2011) and the ways it is defined have become increasingly diverse and context dependent. Stahl et al (2013 p.200) for example considers RRI to be 'a social construct of ascription that defines entities and relationships between them in such a way that the outcome[s] of research and innovation processes lead to socially desirable consequences and importantly, socially desirable for whom and why' and focuses on society as a whole. Von Schomberg (2012) however, highlights business and economic concerns in defining RRI as ´a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products' (Von Schomberg, 2012, p. 9).

Embedding RRI however, requires the provision of tools and guidance, which will only be used and therefore useful if they fulfil the needs of the stakeholders being expected to implement them. Further, embedded ways of working or approaches within research and innovation culture may also be changed through education, and evidence that RRI actually improves outcomes. This means that any guidelines or recommendations, particularly if they require changes in already established working practices, policies and procedures, should contain only that which is needed, workable, relevant and practical and which provide evidence that it will lead to improvements. Guidelines therefore should allow each stakeholder group to develop their own suitable strategies of responsible innovation during all phases of the project life cycle, from planning and implementation to evaluation and revision.

To create such an important and potentially far-reaching document however, first involved understanding what the core principles of RRI are. To this aim, Pelle and Reber (2013) identify the five key ingredients of RRI as:

Anticipation:

In the context of technological development, anticipation tries to predict possible social outcomes by developing scenarios and reflecting on the ethical issues to 'reveal visions of the world associated with a given technology' (Grimpe et al, 2014)

Transparency:

This means that once possible outcomes have been identified, including both desirable and undesirable ones, they should be disseminated and made available.

Responsiveness:

To be responsive in any research and innovation process requires a deliberate reflection on current practices and behaviour. Beyond this there is also a need to adapt and change, not just once, but possibly many times during the course of a project.

Reflexivity:

Two orders of reflexivity provide key ingredients for successful RRI. The first is to consider the extent that something can be adapted or changed in some way so that for example, a problem can be identified and fixed (Pelle and Reber 2014). The second order of reflexivity considers the framing in which the work is done, and whereby researchers and innovators can think about and take responsibility for the assumptions that guide their actions. (Grimpe et al 2014).

Participation:

Participation in RRI is not merely a top-down, tick-box exercise in stakeholder engagement. Participation means that all those affected by or concerned with the process or the outcome of research or innovations should be involved from the outset. (Pelle and Reber 2014, Grimpe 2014).

These key ingredients and definitions of RRI therefore provide underpinnings for the development of the requirements and the subsequent guidelines.

2.2 Norms

It is understood that for guidelines to become normalized in practice, they should be developed in context (Maesschalck and Lenoble 2011) and with an understanding that norms and ways of working may be tacitly embedded and so difficult to identify or change. Understanding the importance of norms in context therefore is a starting point in the identification of the requirements for the guidelines and from which they can also be reviewed and revised. Stahl (2012) explains the need for reflexivity in understanding context, which is important for 'doing' RRI, when he considers that 'engagement with ethical questions will require the development of reflective processes within research, so that norms, their context and application can be understood, predicted and influenced'(Stahl 2012 p.209).

Therefore, to ensure that the reflective process was engaged with by the participants, they were asked to evaluate the requirements for guidelines from two perspectives. Firstly from a professional (institutional, organisational, academic field etc) context and secondly from their own personal (social, ethical, individual) context. In this way both first (problem identifying and solving) and second (norm and context framing) order of reflexivity on the guidelines was achieved. This helped to understand what was important to the stakeholders, what guidelines would mean to them in their personal and professional context, and how those expectations and concerns could provide insight into how to design a set of guidelines for RRI that could actually be used in practice

However, just a reflexive approach to context alone cannot provide answers to what is required in a set of guidelines. A concept of RRI is taking personal responsibility (Owen et al, 2013, Sutcliffe 2011, Fedor et al 2009) alongside an understanding that there may be a disconnection between organizational norms and an individual's normative horizons within their own 'personal' context. This can lead to irresponsible behaviour that whilst generally unacceptable to the individual, is considered an acceptable norm in particular contexts such as in the work-place, where it is 'one thing for a norm …to be acceptable in principle, another… to be valid in practice." (Maesschalck, 2001, p. 83).

An example of a context specific norm of personal responsibility is file sharing online. It is not unusal for individuals to consider the general principle of theft to be wrong, and yet have few qualms about the downloading and distribution of copyrighted material. The changing nature of what constitutes property and therefore theft, has left both ethical and policy vacuums (Moor 1985). Understanding and factoring-in context therefore is a key requirement for influencing change in behaviour, particularly if that is then to become the norm for that individual. In practice, this means that one of the building blocks in devising the requirements for guidelines was that they should be context-specific and support the building of new norms.

It is also understood that there are different approaches to the governance of research and innovation, from the researchers themselves, within organisations and to funding organisations such as the European Commission. Therefore a multi-disciplinary approach is needed to gain insights into established norms within the researcher community and help to understand perceptions and practice of governance within their own context. Through a survey, case study analysis, workshops, semi-structured interviews and focus groups, the findings from this research will be the main sources of evidence for the content in the guidelines. However, this approach could also be applied for other projects.

2.3 Governance and the Participatory Approach

The goal for a set of guidelines for RRI therefore is the effective governance of research and innovation practices leading to a change in behaviour and establishment of new norms in context that reflect RRI principles. This is an ambitious goal and can only be tested in practice. Governance has been identified as 'an attempt to answer a "trilemma" between "scientific accuracy, policy effectiveness and political legitimacy" (Pellizzoni, 2004), i.e. between the rules of scientific knowledge, the efficiency of political norms and rules, and their social acceptability. Further, governance is seen as also being reflexive, again taking context and norms into consideration.

Governance is also self-determining and considers the needs and inter-relationships between the affected actors and tries to envisage the most appropriate course of action.

Governance is often seen as reflexive and self-determining and should consider the needs, relationships and context of those affected (Jessop 2003). Further, given that there are many different ways of conducting and governing research and innovation (Groves 2006) and that these are also likely to be in a range of different contexts, it was understood that a requirement of the guidelines was that they need to be designed in a way that they support different stakeholders' own initiatives within their own context and through a democratic participatory approach (Lenoble and Maesschalck 2003). Governance then, when considered in light of the development of guidelines, requires that decisions are not so much dictated from above by the imposition of one set of rules for all, but that RRI governance should emerge from a more democratic and inclusive process.

The participatory approach (Rowe and Frewer 2000) and concepts of procedural justice, which provides a theoretical perspective on the practical experiences of science policy and the importance of stakeholder involvement in effective decision-making (Joss and Browlea 1999) indicate the importance of democratic ideals surrounding science and technology policy. Democratic approaches to participation (focus groups, workshops, questionnaires and so on) can facilitate acceptance (albeit with limitations) (Jessop 2003) alongside the participatory approach.

However, it is important to avoid public engagement for its own sake, and to avoid the 'de-mocratising of democracy' (Felt and Fochler 2010 p.18). The danger of paying mere lip-service to stakeholder involvement in the process of developing the requirements for the guidelines would mean that any resulting requirements would be unlikely to lead to the development of a set of guidelines that would be acceptable to the stakeholders themselves and would therefore be entirely ineffective.

Awareness of this meant that efforts were made to ensure that the stakeholder views were used to directly inform the content of the requirements, and that the views of each individual were considered of equal weight. The empowerment of the actors through the use of unambiguous, effective and usable guidelines, developed in context and with stakeholder participation makes it more likely that the guidelines will be seen as an enhancement to working practices and lead to embedding RRI governance into research and innovation working behaviour.

However, engagement is just one of the conditions for RRI and the requirements for the guidelines therefore, were also built on an understanding that successful RRI, and in particular any guidelines promoting RRI approaches 'represents the attempt to provide an answer to the multitude of ethical, moral, legal and other problems arising from the use of technology research and innovation' (Von Schomberg in Stahl 2011).
Ideally then the process for creating a set of guidelines should both acknowledge the importance that the role of actors and stakeholders have in establishing their own norms, and consider the many factors and issues that may arise.

3 METHODOLOGY

Having established the need for an inclusive, democratic, reflexive and participatory approach that acknowledges norms in context, it was necessary to provide an initial set of requirements

for guidelines to enable the participative process to begin. To avoid re-inventing the wheel it was decided to utilize existing sources to inform the starting point for the creation of the requirements. A recently completed EC FP7 project, CONSIDER (Civil Society Organisations In Designing Research Governance) had created a set of stakeholder specific RRI guidelines for engagement with Civil Society organizations in research. In addition, FRRIICT (Framework for Responsible Research and Innovation in ICT), another recently completed RRI project funded by the EPSRC (Engineering and Physical Sciences Research Council) had created a framework and tools for RRI in ICT. These projects and the expertise of the GREAT consortium informed the development of the initial set of 14 requirements.

The next stage was to invite a range of stakeholders drawn from researchers across a range of disciplines to a workshop to reflect on what was being created, why it was important and to be engaged with the creation process. This was so that the initial draft requirements could be revised or re-written if necessary to better reflect the needs of the stakeholders. This approach is important to ensure that the process of identifying the requirements for guidelines was not only looking towards providing guidance for future RRI governance, but to also ensure that RRI principles were embedded within its own creation.

During the process, ongoing research was able to further directly inform the theoretical landscape of RRI and the context in which the guidelines were to be produced. Therefore, the stakeholder revised requirements were then further tested through evaluation by the project partners in the light of their own research and experience. In addition, the literature on the approach to the creation of guidelines and frameworks for RRI was further examined to inform their development.

3.1 The Workshop

The stakeholder engagement activity for revising the requirements for the guidelines was selected on the basis that it would enable discourse between actors with coinciding and yet also very different approaches to research and innovation. With one of the core stipulations that the guidelines should address all stakeholder groups, the involvement of people from a range of disciplines, all of whom could be directly affected by RRI guidelines was considered to be important to provide valuable insight.

The workshop itself was approached and conducted in a similar way to a focus group, i.e. problem-centered group discussions moderated by the researcher (Krueger, R & M.A Casey, 2000). In this instance, the discussion centered on the initial set of 14 requirements, as the workshop's intention was to evaluate and provide feedback and suggestions on these initial requirements. Participants were encouraged to reflect on each of the draft requirements and to offer alternative or additional requirements. In this way it was anticipated that acceptance of the resulting guidelines would be encouraged when identified with their own experiences, within their own context, and with acknowledgment of the norms of research and innovation practices within their discipline.

The workshop also encouraged the stakeholders to engage in second order reflexive thinking throughout to 'think about their own ethical, political or social assumptions underlying and shaping their roles and responsibilities in research and innovation as well as in public dialogue' (Pelle and Reber 2014 p.17).

During the course of the workshop, each of the draft requirements was evaluated in turn, to systematically evaluate each one in depth. In addition, the principle of having guidelines for RRI governance, the need and likelihood of acceptance was discussed. This provided significant insight into the perceptions of researchers towards future guidelines for RRI. Whilst this was not the focus of the workshop, the generally dismissive approach to the idea of guidelines in any form merely served to highlight the need to not just impose guidelines, but to facilitate their acceptance through democratic participative approaches and to educate future generations of researchers in the principles of RRI to foster new norms of behaviour.

3.1.1 Participants

In order to effectively and appropriately evaluate the requirements for guidelines, it was important that those invited to participate in the workshop were those stakeholders most likely to be affected by the introduction of guidelines for the governance of RRI. The rationale for selection of the participants in the workshop therefore was based on an understanding that there are multiple possibilities when identifying and selecting stakeholders, some of whom may also have incompatible interests (Friedman and Miles 2006).

The stakeholders invited to participate in the workshop were drawn from those people who were amongst the potential users of the guidelines and thus were considered to have an interest in both their design and development. However, this pool of potential participants is vast and so a further narrowing of potential participants was necessary. In order to select which particular stakeholder groups to focus on, selection utilised criteria that was specifically devised within the project to ensure consistency. However, it is acknowledged that when selecting participants in other projects, the criteria used would be specific to that particular project's needs. In light of this, the participants for this particular workshop were selected from one of the categories below:

- The participants are conducting international research ('cross nation')
- They work in different disciplines or on different research topics
- Technology or management may play a role in the research:
- The expected outcome of the participants' research is a technology, management process or are technological procedures, that may be considered innovative;
- The research process itself involves technological components, management processes or technological procedures that may be considered innovative;
- Information and communication technologies (ICTs) are strong enablers for the scientific research.
- The innovation process, or the expected outcome, involves some risk or uncertainty.
- The participants are at different stages of their academic career (e.g. doctoral student; postdoctoral researcher; professor).

The selection process and subsequent invitations led to seven researchers agreeing to take part. The participants came from a range of disciplines including management, technology, and computer ethics and included:

- two Professors currently involved in European FP7 Projects

- one Postdoctoral / Research Associate involved in a UK based project and an European FP7 project
- one PhD student involved an a UK based project
- three Senior Lecturers/ Senior Research-fellows involved in several European FP7 projects

Of these participants, the Postdoctoral/Research Associate and one Senior Lecturer/Senior Research-fellow were in the early stages of their careers. The other participants were in mid-career stage and one senior stage.

3.1.2 Workshop Structure

In order to allow time for preparation, the workshop participants were sent a participant information sheet and consent form. The information sheet provided an overview of the project, an explanation of what the workshop was hoping to achieve and the initial table of 14 requirements.

At the start of the workshop, there was an introduction to the project, and specifically the requirements for guidelines. Then all participants were asked to provide a brief introduction to their work and to indicate what kind of projects they had worked on or were working on currently. This provided the participants with a clear impression of what was expected from them and to understand some of the different perspectives and approaches of their fellow participants.

There was then a brief discussion of the initial requirements amongst all participants to discuss what they are intended to be used for and what the first impressions were. This was followed by a point by point analysis and evaluation of each element of the initial requirements table. Suggestions for improvement and revision of the requirements were suggested and noted. The workshop was sound recorded and had a note-taker. Whilst there were some extremely valuable suggestions made during the workshop, it was felt that subsequent reflection by the participants could result in further revisions. Therefore, a second revised table, based on the findings from the workshop was sent to all participants to ask for further feedback. There were no responses to this request and the requirements table was then sent to the project partners to enable them to further inform the identification of the requirements for guidelines from their own research and expertise. It was acknowledged that the project partners would also be impacted by the guidelines subsequently constructed based on those requirements.

3.2 The Requirements for Guidelines

The requirements were initially informed by the research findings that led to the first set of 14 requirements for guidelines. It was acknowledged that different stakeholders speak different languages (national; technical; domain-specific), and that most of them have little time and are busy with various tasks. Therefore any further imposition of a new set of regulations on top of already existing ones would not be well received, perhaps seen as further restricting their ability to undertake the actual work. However, adding a further layer of regulation is not what is intended by the guidelines. On the contrary, the intention is for them to be used as a guide for people to better understand how to be responsive and responsible from an ethical perspective and not a legal one which is sometimes seen as box ticking compliance rather than an opportunity to reflect on current practices.

The final 11 requirements detailed below (Wilford et al 2014) were the result of both the initial identification of the requirements for guidelines discussed above, and the subsequent stakeholder engagement process which directly informed the revision of the initial set.

The final requirements are presented in two sections; firstly a set of constructive, process focused requirements were identified. These would indicate the look and feel of the guidelines to make them accessible and usable. Secondly a set of substantive, content focused requirements that would be practical and effective were defined.

3.2.1 Constructive, process focused requirements

1. *Use a common language that overlaps all disciplines.*

One of the challenges for the creation of guidelines is that across different disciplines as well as in different countries, there would likely be language that would be understood in a very specific and contextual way by specific stakeholders. These may be technical terms that would be important to be used for clarification or succinctness, or terms that may have different meanings depending on context. Therefore, it was indicated that where special terms were needed for clarity, a link should be provided to an appendix or website which should include a glossary providing definitions of terms used in the guidelines that would provide consistency in the understanding of what a particular term means in the context of the guidelines.

2. *Be concise and ensure it is practical and usable (bullet points etc.) as shorter documents are more likely to be read and understood.*

As indicated above, researchers often already need to read many documents on a daily basis and so the addition of further 'work' needs to be considered sensitively.

3. *Use good style to enhance readability (colours, diagrams, pictures, other types of media). Make it attractive and easy to understand.*

It is important in a guidelines handbook for RRI governance that it is presented in a way that makes the information easy to understand and to use. The inclusion of graphics and other media means that the guidelines will be accessible to different types of learners (See Gardner 1983 for an in-depth understanding of approaches to learning). In addition, the use of different approaches to present the guidelines will prevent the document from being a purely text based which may not be appropriate for all of the target audience, or may even be off-putting for some users.

4. *Provide an interactive document (e.g. links to RRI websites, case studies, providing examples of 'good'/'bad' practice or normative dilemmas, tools and resources). to provide examples for discussion leading to organisational/individual learning.*

It was felt that a digital interactive 'document' may be more effective and appealing to some stakeholders than purely paper guidelines, particularly with the increasing use of electronic devices such as tablets and mobile phones to access information.

By providing the information electronically and online, the ability to link directly to the glossary and other resources will enable decision-makers to better contextualize their own RRI approach.

5. *Provide a pitch to grab attention, for example, a cover page with the key points.*

A document that is 'eye-catching' is more likely to be actually picked up and read. In addition, a casual observer may also be attracted to such a document, thereby encouraging further dissemination of the message of RRI beyond the core target group of researchers and innovators.

3.2.2 Substantive, content focused requirements

6. *Provide a small number of concise RRI definitions and other key terms that are tightly coupled to the findings from the project.*

There are a host of definitions of RRI that provide context and discipline specific focus. In developing their own approaches to RRI and to facilitate the development their own RRI approach, multiple definitions may be needed. However, detailed definitions and their explanations may conflict with requirement 2 *(Be concise and ensure it is practical and usable)* to be concise. Therefore, within the guidelines themselves, only a small number of selected general definitions should be offered. The provision of external links to other definitions will enable wider interpretations of RRI to be considered if needed.

7. *Provide links to further definitions of RRI to broaden awareness of RRI principles and to encourage the use of RRI theory to relate to user's own practice.*

The links to definitions and other resources would be provided to help researchers to identify the scope of RRI and the importance of embedding its practices within their own research and innovation context. This will also go some way to avoid tick-boxes and bolted on practices. This is made more likely if just one approach or perspective is offered and could then limit the amount of change possible within a particular discipline or organization.

8. *Provide methods to re-asses and challenge the guidelines including reflection on the processes,outcomes and impact of the guidelines*

Research and innovation is by its nature dynamic and ever changing. It is therefore essential that any guidelines for an RRI approach should be under regular review, partly to reflect the flexibility of the guidelines, and partly to ensure that practical relevance is maintained.

9. *Respond to the EC framework, e.g. intervention logic model (relevance, effectiveness, efficiency, utility) and relate the benefits and problems of RRI to the EC framework.*

In view that the guidelines created in this research are largely aimed at EC researchers, it was felt that they should identify and respond to aspects of the European Commission framework that are specific to the project, area of enquiry, stakeholder group etc. It is understood however that in some cases this may be too prescriptive, narrow in scope or it may not be accepted in other geographical regions and may create confusion where there are conflicting demands. Where this occurs, then it was felt that legal requirements should take precedence.

10. *If the pluralistic approach to RRI currently developed in the project turns out to go beyond the scope of requirement 6 ('provide only a small number of concise RRI definitions'), deliberate on possible ways of representing this pluralistic approach without compromising too much on requirement 2 (Be concise and ensure it is practical and usable.)*

This requirement further encourages flexibility within the guidelines and the ongoing discourse on RRI and how they can be presented. As technology and expectations change, the approach to the presentation of future iterations of the guidelines needs to be under regular review. A more pluralistic approach may require even greater need for flexibility in the guidelines and revisions would need to reflect this.

11. *If explicit norms of responsible behaviour are expressed in the guidelines, these norms should be established with the participation of stakeholders, and by taking into account their contexts.*

This requirement rests on one of the key findings of the project in that 'good' governance implies, among other things, that various actors participate in the making of the very norms they subsequently have to follow. In this way, the resulting guidelines will aim to help to establish new norms of behaviour and to facilitate the normalisation of responsible research and innovation practices into the future.

4 CONCLUSION

The guidelines for the governance of RRI can be directly informed by the requirements identified through the process detailed above. However, the flexibility of RRI means that these steps can also be applicable to other projects where the development of guidelines are required. It is understook that there is always be scope to gain further understanding about what is needed in requirements and subsequent guidelines and so future guideline development and understanding of requirements would be enhanced through being applied in other projects and areas of inquiry.

Further, and in keeping with the five principles of RRI, it is *anticipated* that any requirements identified are likely to change over time; the process should be *transparent* in the way that changes are introduced and *responsive* to the needs of society as well as funding bodies, scientists and researchers. In addition the importance of being *reflexive,* not only about the processes and

procedures, wider impact, and unexpected consequences of those actions but also to consider the framing of the requirements and the norms in context from the personal perspectives of the stakeholders.

Finally, should the guidelines in practice or the rationale for the requirements change so that revision is needed, or if current practice either directly or indirectly causes harm, then, in particular, the *participation* of those affected should be prioritized to ensure that changes made are also decided through utilising an RRI approach.

Throughout the process of the identification of requirements for guidelines, and subsequent guidelines derived from them, it was important ot emphasise that they should not only incorporate RRI principles into the guidelines themselves, but they should also construct them incorporating RRI principles. In this way, the perception of legitimacy of both RRI and the resulting guidelines is reinforced and the applicability of the process to the development of guidelines in other areas is strengthened.

5 ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's Seventh Framework Pro-gramme (FP7/2007-2013) under grant agreements n° 321480 (GREAT).

Thanks to the GREAT project consortium for their valuable input to the process of developing the requirements for guidelines.

6 REFERENCES

[1] EPSRC (2014) *Framework for Responsible Innovation* http://www.epsrc.ac.uk/index.cfm/research/framework/

[2] Fedor, Carol. A, Cola, Philip A, Pierre, Christine (Eds) (2006) *Responsible Research: A Guide for Coordinators.* Remedica.

[3] Felt, U., and Fochler. M., (2010) 'Machineries for Making Publics: Inscribing and De-Scribing Publics in Public Engagement.' https://sts.univie.ac.at/fileadmin/user_upload/dep_sciencestudies/pdf_files/Preprints/felt_fochler_machineries_Aug2010.pdf. Accessed 15/10/14

[4] Gardner, H. (1983). *Frames of Mind.* New York: Basic Book Inc.

[5] GREAT (Governance of Responsible Innovation) FP7 Grant Agreement No: n°321480 http://www.great-project.eu/

[6] Grimpe, B., Patel, M., Jirotka, M., Wilford, S., Niemelä, M., Ikonen, V (2014) 'Context of RRI report' *GREAT (Governance of Responsible Innovation)* FP7 Grant Agreement No: n°321480

[7] Pelle, S. and Reber, B (2013) 'The Theoretical Landscape' *GREAT (Governance of Responsible Innovation)* FP7 Grant Agreement No: n°321480 http://www.great-project.eu/deliverables_files/deliverables03

[8] Groves, C. (2006) 'Technological Futures and Non---Reciprocal Responsibility' *The International Journal of the Humanities,* 4 (2), pp. 57-61.

[9] Jessop,B. (2003) 'Governance and Metagovernance: On Reflexivity, Requisite Variety, and Requisite Irony', in H. P. Bang (ed.), *Governance, as Social and Political Communication.* Manchester, UK: Manchester University Press, pp. 142-172.

[10] Joss S. and Browlea A., (1999) 'Considering the Concept of Procedural Justice for Public Policy – and Decision-Making in Science and Technology dossier'. 'Special Issue on Public Participation in Science and Technology', *Science and Public Policy*. Vol. 26, N° 5, October 1999, pp. 321-330.

[11] Lenoble, J., Maesschalck, M. (2003) *Toward a Theory of Governance: The Action of Norms*. New York: Kluwer Law International.

[12] Maesschalck., M.(2001) *Normes et Contextes*. Hildesheim: Georg Olms Verlag.

[13] Moor, J. (1985) What is computer ethics?*Metaphilosophy.* Blackwell.

[14] Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*. *39*(6), 751–760. doi:10.1093/scipol/scs093

[15] Pellizzoni., L. (2004) 'Responsibility and Environmental Governance' *Environmental Politics*.

[16] Rowe G. and Frewer J.L., (2000) 'Public Participation Methods: A Framework for Evaluation.' *Science,Technology, & Human Values*. Vol. 25, N° 1, Hiver pp. 3-29.

[17] Sandri, S (2009) *Reflexivity in Economics: An Experimental Examination on the Self-Referentiality of Economic Theories.* Springer-Verlag Berlin Heidelberg 2009

[18] Stahl, B. C and Wakunuma, K (2015) *Guidelines Handbook. CONSIDER (Civil Society Organisations in Designing Research Governance.)* FP7 Grant Agreement No: 288928 http://www.consider-project.eu/

[19] Stahl, B. C., Eden, G., & Jirotka, M. (2013). Responsible Research and Innovation in Information and Communication Technology. Identifying and engaging with the ethical implications of ICTs. In R. Owen, M. Heintz, & J. Bessant (Eds.), *Responsible Innovation*. Chichester, UK: Wiley.

[20] Stahl, B. C. (2012) 'Responsible research and innovation in information systems' *European Journal of Information Systems.* (2012) 21, 207–211

[21] Stilgoe, J., Own, R., and Macnaghten, P. (2013) 'Developing a framework for responsible innovation' *Research Policy*. Volume 42, Issue 9, November 2013, Pages 1568–1580

[22] Sutcliffe, H. (2011). *A report on responsible research & innovation. DG Research and Innovation of the European Commission.* Located at Http://ec. Europa. Eu/research/science-society/document_library/pdf_06/rri-Report-Hilarysutcliffe_en. Pdf. Retrieved from http://www.matterforall.org/pdf/RRI-Report2.pdf

[23] Von Schomberg, R. (2012). Prospects for Technology Assessment in a Framework of responsible research and Innovation. In M. Dusseldorp & R. Beecroft (Eds.), *Technikfolgen abschätzen lehren.* (pp. 39–61). VS Verlag für Sozialwissenschaften. Retrieved from http://link.springer.com/chapter/10.1007/978-3-531-93468-6_2

[24] Wilford, Sara., Timmermans, J., Grimpe, B., and Jirotka, M (2014) 'Requirements for guidelines' *GREAT (Governance*

of Responsible Innovation) FP7 Grant Agreement No: n°321480

When brain computer interfaces move from research to commercial use

Christian B. J. Hansen
De Montfort University
The Gateway, Leicester
LE1 9BH
+44 7400 714241
christian.hansen@email.dmu.ac.uk

ABSTRACT

This paper will explore how ethical concerns change when brain computer interfaces move from a research setting into a commercial setting. This paper will argue that the transition from research to commercial settings might change the intentions for the artefact and will explore hypothesis of what this change might affect. This paper will discuss how possible intentions for brain computer interfaces in commercial settings will have an impact on the products developed and what consequences this might have for individuals and society. The ethical concerns discussed in this paper includes privacy, enhancement and the digital divide. This paper will also present possible future research which could help investigate both the hypothesis put forward and the topic of brain computer interfaces moving from research to commercial settings in general.

Categories and Subject Descriptors

K.4.1 **[Computers and Society]:** Public Policy Issues – *Ethics, Privacy, regulation*

General Terms

Human Factors

Keywords

Brain Computer Interfaces, BCI, Responsible Research and Innovation, RRI, Ethics, Privacy, Enhancement, Equity and Digital Divide.

1. INTRODUCTION

The European Commission has put the agenda of making responsible research and innovation a top priority, which in return have created a lot of focus on how to make research and innovation responsible [20]. While there has been a focus on how to do so, there seems to be a lack of information on what happens to ethical concerns when technology makes the transition from research to commercially available products. Therefore, this paper will explore in which way ethical concerns change for brain computer interfaces (BCI) when making the transition from research to commercial usage. Specifically this paper will focus on how intentions for the BCI will change the product that ends up being developed, and what consequences this will have for the end user and society. It is this focus on the relationship between developers, the BCI and the consequences of these that are interesting in this article.

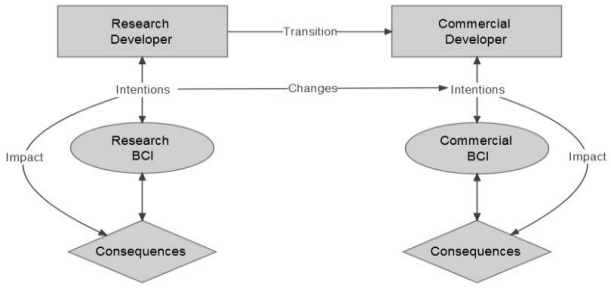


Figure 1: Diagram of change from research to commercial development

This paper will explore the current literature dealing with the ethics of BCI in research settings, and provide hypothesis of what occur when the intentions for the developed BCI changes. The focus in this paper will be on the change in intentions for the BCI and not on the people or organizations behind this technology. The change in intentions for the artefact, changes the impact BCI will make on individuals and society. Therefore by exploring how the intention of BCI changes the consequences it might have, we can explore what further research could answer the questions that arise from this change.

2. BACKGROUND

This section will describe two different discourses, specifically the discourse of ethics of neuroscience [23] which will be referred to as neuroethics and the discourse of responsible research and innovation (RRI) [25], however firstly a short description of BCI will be made.

Brain computer interfaces take many forms, such as invasive, non-invasive, wet and dry BCI. An invasive BCI is a BCI that uses interfaces that are implanted directly onto the brain. These devices are rarely used by healthy individuals as these require surgery and are not biologically sustainable which introduce a lot of risk factors. A non-invasive BCI is a technology that reads signals on the surface of the scalp instead of directly reading signals from the brain. This technology has more potential usage for healthy individuals as these can be compared to wearing a smart watch, using a keyboard or even a computer mouse [16]. This paper will be focusing on specifically non-invasive BCI, as non-invasive BCI are now emerging as commercial products. This paper will

particularly focus on dry BCI as these are the easiest of the wide variation of BCI to commercialize as they require the least preparation from users and are non-invasive. Dry BCI differ from wet BCI by not using any gel which reduces the preparation time, in return their accuracy and ability to read signals are reduced [16]. There is also good reasons to believe that particularly dry non-invasive BCIs will be the most prominent market of BCIs in the near future. This is based on the BNCI Horizon reports, which report of 51% of industries surveyed using some sort of electroencephalography (EEG), and only 6% of industries using invasive BCI. Current BCIs work by reading the electric activity across the scalp. By doing so it is possible to create a model of where activity is present in the brain. Various techniques are then used to provide meaning to this data, such as algorithms that measure the difference between a resting state and an active state to provide either actions or feedback based upon these two states. Companies are using this technology to provide users with a commercial product they can use for various tasks such as therapy (including meditation), entertainment or research [3, 10].

The history of neuroethics has dealt with ethical concerns regarding research into the brain and brain computer interfaces. Topics that has been dealt with range from privacy, to enhancement, [14, 18, 29] however there is a lack of research looking at the different stages of neurotechnology (such as brain computer interfaces) development, and what the different stages might have of impact on the ethical concerns. Specifically neuroethics have discussed privacy, both where it is suggested that the information collected is not different than the information collected in psychological research [1]. Others point at the predicting nature of neuroimaging as similar to genetic information and suggest that the same privacy laws used to regulate genetic information is used as basis for laws to cover neuroimaging data. [24] Whether or not the information is classified at the same privacy level of genetic information, information collected by these devices is none the less private and therefore must be treated as such. What both sides in this debate does not cover is what the change from a research setting to a commercial setting means to privacy concerns. Therefore in dealing with privacy concerns the current literature discuss potential issues, or issues related to keeping data private in a research setting. But by doing so the issue of privacy is focused around a research setting, a commercial setting or both. This however leads to a gap in knowledge of what is changing with the potential ethical concerns when new steps in the development are taken. This gap in knowledge is however not isolated to privacy concerns, but is the case for most ethical concerns.

The discourse of responsible research and innovation is mainly focused on the development and specification of what RRI is, or how RRI can make an impact on current research and industry. This can be attributed to the fact that RRI is a fairly new term. While RRI is a new term it leans upon discourses that are more settled such as technology assessment and computer ethics. While these discussions are interesting they leave out an important element, which is the differences between research, and commercial innovation. When the discourse is interested in the differences between research and commercial settings, it is largely on how research can be adopted in commercial settings and society [13, 15]. By not focusing on this difference there is a risk of missing out important aspects of technology assessment and computer ethics. So while good effort is being put into asking questions on how to impact research and innovation, there seems to be a lack of discussion on what the differences are between commercial innovation and research. Due to this lack of focus on the topic, areas which are specific to either innovation or research might be missed.

3. INTENTION

While intentions of humans are a philosophical topic that has been discussed in great effort [8, 21, 27], in this article intentions are meant as for what purpose a BCI is designed and the intended use for the BCI. It is in this context intentions are to be understood throughout this article. As in figure 1 in the introduction, the intentions we are talking about here is the relationship between the BCI developer and the product that is interesting. Therefore the intentions discussed are the intentions for the artefact, and not the intentions of the artefact or the developer of the artefact. If the intentions for the artefact is to provide wheelchair users with another interface to control their wheelchair, the device will look differently than if it were designed for playing a video game. It is these changes in focus of attributes that this paper will focus on and discuss what these different focuses might mean for the ethical consequences for individual users and society. One could argue that the actual consequences of an artefact might not be predictable and possible to design for or against. While this is true, it is still an interesting exercise to speculate what might happen when technology is moved from research to commercial use. Therefore in the following sections, the consequences brought forth might not be complete and there is good reason to believe new concerns will emerge once this change is complete. The reason why intentions are worth looking at is because of the addition of values to artefacts change based on the intentions for the artefact, such as described by literature in value based design [12, 19]. For example the Intel-chip case brought up by Nissenbaum (2001) shows that the intentions for making a chip more secure and protected against hardware theft raised privacy concerns. In a similar way the intentions for a commercial BCI might have other ethical concerns than a BCI developed for research. It is this change in intentions and thereby consequences the following sections will discuss.

4. CONSEQUENCES

The following sections will discuss how the values and intentions embedded into a commercial brain computer interface might affect the consequences to individuals and society compared to BCI developed in a research setting. This section will discuss the ethical concerns of privacy, enhancement, and the digital divide. While these ethical concerns are not a comprehensive list of concerns, it is some of the concerns which are likely to be raised when brain computer interfaces move from research to commercial settings. As mentioned previously, the consequences mentioned in the following sections are hypothesis of might happen, and future research will be required to evaluate whether these hypothesis are true.

4.1 Privacy

In research, privacy might not always be a value directly in focus when developing BCI products because privacy in part is handled by organizational protocols such as ethical reviews, restrictions to ownership of data, and other means of protecting users and society from data gathered to be misused or disclosed. When the technology moves into commercial usage there is various

interesting possibilities for the value of privacy in the brain computer interface, and this paper will hypothesise on three of these possibilities. The three possibilities discussed will be, developing a BCI with privacy in mind, with share-ability in mind, and finally developing without privacy or lack thereof in mind.

If the value of privacy is being embedded into the BCI the consequences could be that for those individual users that value privacy would be more likely to adopt the technology. The same privacy could however mean that for society there is less ability for policing what data is being collected and for what purpose as it would be harder to gain insight. The worry of not having easy access to user data generated with BCI devices should not be as much as a worry though because the data which should be interesting for law enforcement is data that is already collectable. This is data that is closer to output rather than input of the BCI, just as keyboard inputs are interesting, but not whether the user is typing with fingers or another limb. One could argue that law enforcement would be interested in direct BCI inputs in form of brain wave data as it could be used to identify certain states of mind. This is however still a future scenario as research at this point is not able to use BCI data in such a way. This is still an interesting topic, which fortunately is already ongoing, and something we as society need to take a stance on [24, 26]. It might also be that while more products would be sold, there is less options for companies to make a profit as there is less options for using user data as a product. This could slow down commercialization as there would be less incentives for companies to develop BCI products.

If however the intention of the BCI is to make a product where the BCI itself is not the main source of income. We could see BCI products that focus on gathering user data, and sharing these with commercial third party companies. This could increase the amount of BCIs sold as it would make it possible to sell products cheaper, however it would in return raise issues of privacy. Whether people would be willing to give up their rights to information about what is going on in their brain is a question worth asking. Companies such as Google and Facebook have been successful in providing products to the world in exchange for personal data, whether this will be a potential business model for BCI developers is yet to be seen. The consequences for BCI products if this course of development is chosen would be that more people would be able to gain access to these devices which could be argued as a good thing. The question remains however whether too much privacy is traded for a cheaper product. Commercial products such as those provided by Facebook seem to have made the concept of privacy fussy in regards to digital privacy [30]. The same sort of change as seen in digital privacy could be an impact of BCI to the concept of brain activity being private.

Lastly it might be that there is not going to be any focus on privacy in the BCI device. This might be the most difficult to predict as this leaves out both possibilities and could lead to devices that have privacy concerns without it being the intent. Therefore having a position on whether the BCI development should be developed with the intent of being privacy enhancing or not is important as it at least forces commercial developers to take a stance on what they want their devices to be used for.

4.2 Enhancement

The consequences for enhancement when the intentions for the BCI is commercial viability is very hard to evaluate as there is many different notions of what the concept of enhancement covers. The change in effect however might be the most noticeable to society as there is large potentials for companies to reach a large number of people. BNCI Horizon (2015) mentions that there is over 100 million students in the EU alone, and even with just a percentage of these students using BCI would be a large market to reach into [22]. The ethical concerns with this type of introduction to enhancement is that there is no oversight in both the way people enhance themselves when it is a commercial product, and there is no oversight to who is able to access these devices. This creates a large set of issues that also were a concern in research, but these issues were regulated just like with privacy concern. The major change here is that while enhancement BCI in research settings will be focused on gathering new knowledge and progressing research, enhancement BCI in a commercial setting would be focused on marketability. In a research setting having highly accurate results would be a major concern, whereas this might not be as much as a concern for commercial products as long as the results that were provided could be marketable. The major issue at hand regarding enhancement will be how BCI will be defined. There is two discourses which BCI could follow in this definition, which is either as a training device, or an enhancement device. If BCI is defined as a training device, the effects of a BCI would be categorised as the effects of a treadmills effect on muscle development. If this definition is used one could argue that there is no ethical concerns in regard to enhancement as the ethical concerns regarding digital divide could be solved in the same way as with physical training devices such as treadmills and exercise bikes with training centres. Due to the relatively low hardware costs such a solution could be viable and such centres could offer relatively low fees for such a service which would make it possible for most to gain access to this type of training form. Whether this definition is the most appropriate is however still unclear, and further research needs to be made to determine whether this definition is appropriate. If BCI devices are considered enhancement devices it would indicate that the device should be categorized as a medical device which should only be used by trained therapists. This definition would have implications for the current commercial BCI devices as many of these are being sold as self-therapeutic/training devices [3]. They would be under stricter regulations, and this could make it more difficult to make the transition from research to commercial BCI devices. It could increase prices of BCI devices as enhancement devices which would make the issue of digital divide in enhancement technology more of a concern. Another concern with BCI as enhancement or training devices is whether the training or enhancement translates into other tasks. While the effect of BCI training has been documented to be there in different tests, further research needs to be made to investigate whether this effect of BCI training translates into other settings [6, 28]. Therefore further research is required in both possible future enhancement/training settings for BCI to investigate whether the definition of BCI devices should be defined as either enhancement, or training devices.

4.3 Equity and digital divide

Concerns regarding the digital divide have in some way been minimized by BCI technology moving from research to commercial settings. The concern about everyone having access to the technology is being targeted by commercial developers making this technology available to everyone, and not only researchers or specialised technicians. The transition from

research to commercial availability does not solve all digital divide concerns though. Three other concerns about the digital divide is having access to updated technology, motivation for use of technology, and having abilities to use the technology [2, 7, 9]. While having access to technology can be solved by commercial competition leading to lower prices and organisations such as libraries making the technology available, the other concerns are not as easily solved. At the moment the problem of having motivation to use the technology is a focus for both researchers and commercial developers. This is done by looking at usability concerns and making sure there is potential usage for the average user [4, 5, 11]. By doing this the intention for the BCI is to be usable and for the consumers to have motivation to buy and use the BCI. The concern regarding the digital divide and motivation should therefore be focused around motivating those who currently does not have any motivation for using BCI. This seems to be a natural concern for commercial developers however as they always are concerned about trying to get as many users to adopt and use the technology they are selling. The concern about outdated hardware and software is for BCI similar and parallel to the concerns about the digital divide in general. BCI devices could be developed with that concern in mind though, by making the devices more modular and thereby making it easier and cheaper to update selected pieces of the hardware. If a BCI is developed with exchangeable electrodes and components it would mean that BCI users easily could update the electrodes when better electrodes were released, and if a standard of how these electrodes were connected to the main interface, electrodes from different companies could be switched out to provide cheaper alternatives while keeping the hardware updated. The last concern discussed in this paper is the concern of having ability to use the technology. This requires for the before mentioned concerns to be considered and dealt with, as it is difficult to solve this without having users that are motivated to use the technology, and who have access to updated technology. In research this concern in some cases are boiled down to the question of whether a paralyzed patients is able to operate a BCI [17] When it moves into commercial usage, the concern however is how to develop a product that deals with this concern in general. This could be framed as a usability concern and be dealt with in those regards. If a BCI is easy to use, more people will be able to operate a BCI. This does not necessarily solve the digital divide concern though as there will still be a difference between users who have much exposure to the technology and those who is only exposed to the technology in limited ways. If it is possible to develop a BCI with the digital divide in mind though, it could solve some of the digital divide concerns currently existing with human-computer interfaces. If a BCI is to be developed with the intentions of making interaction with computers more intuitive and easier to use than the standard keyboard and mice, it could introduce members of society to technology and information in a way that keyboard and mice could not. By increasing the ways of interacting with technology it would allow for people to seek information and use technology to solve problems and thereby reducing the digital divide. If the intentions for the BCI is to be an enhancement of current interaction such as an addition to the traditional keyboard and mice, it could however mean that the digital divide would increase further as the complexity of human-computer interaction would increase. Thereby the intentions for the BCI in regards to the digital divide could both help reducing the gap, or widening it.

5. Future research

There is a lack of research done on the intentions for specifically commercial BCI technology. Therefore the change in intentions for the BCI is an interesting research topic worth looking into as it might reveal some interesting differences and similarities between the intentions of research and commercial BCI technology.

Future research should focus on answering the questions of what kind of changes stakeholders see as possible changes in ethical concerns, and what makes up for changes in attributes of a commercial BCI and a research BCI. There is also be a need for research into methods of preventing or dealing with ethical concerns in regards to commercial brain computer interfaces.

Currently there is also a lack of knowledge about the intentions for the BCI and how these intentions change the consequences of the BCI development to society and individuals. Once this research has been done, the next step would be to investigate how these concerns relate to each other, and to what degree different stakeholders find them important.

Overall there is a lot of further research to be done in this field, as there is a lack of knowledge in regards to what happens to ethical concerns when the setting changes, but also specifically about brain computer interfaces and their development in both research and commercial settings.

6. Conclusion

In this paper the current discourses in neuroethics and responsible research and innovation were explored. A gap in knowledge regarding technology moving from research to commercial settings were identified. Hypothesis were then explored about what happens when the intentions for brain computer interfaces changes by the transition from research to commercial settings. Specifically topics such as privacy, enhancement and the digital divide were discussed. Hypothesis about what would happen if the intention for BCI were various degrees of privacy enhancing were explored. Concerns about enhancement were explored such as the definition of BCI as enhancement or training technology. The digital divide concern were explored, explicitly concerns about access, skills and motivation were discussed. Considerations of what further research could include were also made, such as research exploring the hypothesis discussed throughout the paper.

7. ACKNOWLEDGMENTS

Thanks to Prof. Bernd Stahl and Dr. Catherine Flick for your support with this paper.

8. REFERENCES

[1] Arstila, V. and Scott, F. 2011. Brain Reading and Mental Privacy. *Trames. Journal of the Humanities and Social Sciences*. 15, 2 (2011), 204–212.

[2] Ball, J.W. 2011. Addressing and overcoming the digital divide in schools. *The health education monograph*. 28, 3 (2011), 56–59.

[3] Biosensor Technology | NeuroSky: *http://neurosky.com/*. Accessed: 2014-11-28.

[4] Bonaci, T., Calo, R. and Chizeck, H.J. 2014. App stores for the brain: Privacy & security in Brain-Computer

Interfaces. *2014 IEEE International Symposium on Ethics in Science, Technology and Engineering* (May 2014), 1–7.

[5] Bos, D.P. 2014. *Improving usability through post-processing*.

[6] Corralejo, R., Member, S., Member, S., Álvarez, D., Hornero, R. and Member, S. 2014. Assessment of Neurofeedback Training by means of Motor Imagery based - BCI for Cognitive Rehabilitation. (2014), 3630–3633.

[7] Crossing the Digital Divide: Bridges and Barriers to Digital Inclusion: 2011. *http://www.edutopia.org/digital-divide-technology-access-inclusion*. Accessed: 2015-06-02.

[8] Davidson, D. 1963. ACTIONS, REASONS, AND CAUSES. *The Journal of Philosophy*. 60, 23 (1963), 685–700.

[9] DiMaggio, P., Hargittati, E., Celeste, C. and Shafer, S. 2004. Digital Inequality: From unequal access to differentiated use. *Social inequality*. K. Neckerman, ed. Russel Sage Foundation. 355–400.

[10] Emotiv | EEG System | Electroencephalography: *http://emotiv.com/*. Accessed: 2014-11-28.

[11] Van Erp, J., Lotte, F. and Tangermann, M. 2012. Brain-Computer Interfaces: Beyond Medical Applications. *Computer*. 45, 4 (2012), 26–34.

[12] Friedman, B. and Kahn, Jr., P.H. 2003. Human Values, Ethics, and Design. *The Human-Computer Interaction Handbook*.

[13] Grimpe, B., Hartswood, M. and Jirotka, M. 2014. Towards a Closer Dialogue between Policy and Practice : Responsible Design in HCI. (2014), 2965–2974.

[14] Haselager, P., Vlek, R., Hill, J. and Nijboer, F. 2009. A note on ethical aspects of BCI. *Neural networks : the official journal of the International Neural Network Society*. 22, 9 (Nov. 2009), 1352–7.

[15] Hempel, L., Ostermeier, L., Schaaf, T. and Vedder, D. 2013. Towards a social impact assessment of security technologies: A bottom-up approach. *Science and Public Policy*. 40, 6 (Dec. 2013), 740–754.

[16] Introduction to Modern Brain-Computer Interface Design: 2013. *https://www.youtube.com/watch?v=Wlwvgm3AHvc&index=1&list=PLbbCsk7MUIGcO_lZMbyymWU2UezVHNaMq*. Accessed: 2015-04-27.

[17] Kübler, a., Nijboer, F., Mellinger, J., Vaughan, T.M., Pawelzik, H., Schalk, G., McFarland, D.J., Birbaumer, N. and Wolpaw, J.R. 2005. Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*. 64, (2005), 1775–1777.

[18] Nijboer, F., Clausen, J., Allison, B.Z. and Haselager, P. 2013. The asilomar survey: Stakeholders' opinions on ethical issues related to brain-computer interfacing. *Neuroethics*. 6, (2013), 541–578.

[19] Nissenbaum, H. 2001. How computer systems embody values. *Computer*. 34, 3 (2001).

[20] Owen, R., Macnaghten, P. and Stilgoe, J. 2012. Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*. 39, 6 (Dec. 2012), 751–760.

[21] Rawls, J. 2012. Two Concepts of Rules. *Interpretation A Journal Of Bible And Theology*. 64, 1 (2012), 3–32.

[22] Roadmap 2020: 2015. *http://bnci-horizon-2020.eu/roadmap*. Accessed: 2015-01-08.

[23] Roskies, A. 2002. Neuroethics for the New Millenium Commentary. *Neuron*. 35, (2002), 21–23.

[24] Safire, W. 2005. Are Your Thoughts Your Own ?: " Neuroprivacy " and the Legal Implications of Brain Imaging The Committee on Science and Law. (2005).

[25] Schomberg, R. von 2013. A Vision of Responsible Research and Innovation. *Responsible Innovation Managing the Responsible Emergence of Science and Innovation in Society: Managing the Responsible Emergence of Science and Innovation in Society*. R. Owen, J. Bessant, and M. Heinzt, eds. Wiley. 51–74.

[26] Schreiber, D. 2012. On social attribution: implications of recent cognitive neuroscience research for race, law, and politics. *Science and engineering ethics*. 18, 3 (Sep. 2012), 557–66.

[27] Searle, J.R. 1980. Minds , brains , and programs. (1980), 417–457.

[28] Toppi, J., Risetti, M., Quitadamo, L.R., Petti, M., Bianchi, L., Salinari, S., Babiloni, F., Cincotti, F., Mattia, D. and Astolfi, L. 2014. Investigating the effects of a sensorimotor rhythm-based BCI training on the cortical activity elicited by mental imagery. *Journal of neural engineering*. 11, (2014), 035010.

[29] Wahlstrom, K. 2013. Privacy and Brain-Computer Interfaces: clarifying the risks. *AiCE 2013* (Melbourne, 2013), 1–8.

[30] West, A., Lewis, J. and Currie, P. 2009. Students' Facebook "friends": public and private spheres. May 2015 (2009), 37–41.

So What If the State Is Monitoring Us?

Snowden's Revelations Have Little Social Impact in Japan

Kiyoshi Murata
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
kmurata@meiji.ac.jp

Yasunori Fukuta
School of Commerce
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
yasufkt@meiji.ac.jp

Yohko Orito
Faculty of Law and Letters
Ehime University
3 Bunkyo-cho, Matsuyama
Ehime 790-8577, Japan
orito.yohko.mm@ehime-u.ac.jp

Andrew A. Adams
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
aaa@meiji.ac.jp

Ana María Lara Palma
Civil Engineering Department
Management Area
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
amlara@ubu.es

ABSTRACT

This study investigates the attitudes towards and social impacts of Edward Snowden's revelations in Japan through a questionnaire survey and follow-up interviews with Japanese youngsters as part of an international cross-cultural analyses. The survey results showed striking contrasts with ones in other countries reflecting the Japanese socio-cultural and political environment.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues: abuse and crime involving computers, privacy, use/abuse of power

General Terms

Security, Human Factors, Legal Aspects

Keywords

Edward Snowden, privacy, state surveillance, social impact, Japan

1. INTRODUCTION

Edward Snowden's revelations about the operations of the US' NSA (and its primary partner the UK'S GCHQ, as well as the intelligence agencies in Canada, Australia and New Zealand) which started on 5th June 2013 confirmed many of the worst fears of privacy/anti-surveillance activists and academics, and even some of those previously dismissed as conspiracy theory nonsense. Both his act of revelation and the activities he exposed have attracted heavy doses of both praise and censure; whereas some have positively evaluated his deed as an act of valour to protect democracy against the tyranny of the state, others have criticised him as a traitor to a country that have been preoccupied with responses to the threat of terrorism since the 9.11 attacks. Indeed, the US government filed charges of spying against him on 21st June 2013, and he has been living in exile in Moscow. He said that only the American people could decide whether sacrificing his lifestyle was worth it by their response [13]. However, it is clear that the issues Snowden raised are not just for American citizens.

Lively discussions of national security, safety and security of societies, personal freedom and privacy have been generated in many countries by Snowden's revelations with many books and papers recently published [13, 7, 8, 9, 25, 26]. However, there seem to be differences in press and government reactions, so the question of ordinary people's attitudes towards the revelations, and, therefore, their social impacts in different social contexts [1]. This study deals with the attitudes and social impacts in Japan, taking the Japanese socio-cultural and political environment into account. The survey was first developed in English as a base-line and then translated into other languages for deployment in various countries. The study in Japan was conducted in Japanese although the original English versions of the questions and answers are presented here.

We begin by describing the background of government surveillance in Japan in the modern era (post-1868), split into pre- and post-WWII. Following that, we describe the survey presented in this paper and its relationship to a broad international deployment of the same survey in multiple other

countries. Next, we present the results of the analysis of survey responses in Japan. We finish with some concluding remarks about the particulars of the Japanese responses, and some brief comparisons with the results of the survey in other countries.

2. STATE SURVEILLANCE IN MODERN JAPAN

2.1 Before the Defeat in the Asia-Pacific War

Since the restoration of the monarchy in 1868, Japan's governments have continually focussed on preservation of a stable national polity, embodied in Articles 1 and 4 of the Constitution of the Empire of Japan (February 1889) [10]:

> Article 1. The Empire of Japan shall be reigned over and governed by a line of Emperors unbroken for ages eternal.
>
> Article 4. The Emperor is the head of the Empire, combining in Himself the rights of sovereignty, and exercises them, according to the provisions of the present Constitution.

This initial political framework centred on the emperor's ruling power [15], a principle affirmed by the Supreme Court in May 1929 [16]. In order to maintain that principle, the Home Ministry played a pivotal role in surveilling and disrupting the activities of anyone suspected of wishing to disrupt that order.

After the end of the Seinan War in 1877 (the last inland war in Japan) the Freedom and People's Rights Movement demanded establishment of a constitution and national diet which would guarantee freedom of speech, movement and assembly. This movement was the target of and impetus for the creation of the Higher Police Division and their extensive state surveillance operations [20].

The original oligarchy surrounding the Meiji Emperor was supplanted by the establishment of the Imperial Diet in 1890 including the development of party politics and the emergence of a political cabinet in 1898 shifted the focus of government surveillance away from this civil rights group, which had largely achieved their goals despite it. On the other hand, the rapid industrialisation and transition to a capitalist economy in Japan in the late 19th and early 20th centuries led to the emergence of other groups challenging the central government's power. Issues such as growing industrial pollution, typified by tne Ashio Copper-mine poisoning incidents [27] led to the development of social, labour, agrarian and socialist movements. Those engaged in these movements were regarded as a threat to the national polity and so targets of state surveillance. The Public Order and Police Law (Act No. 36 of 10 March 1900) regulated citizen's (subjects') political activities and was used to crack down on these movements.

In the wake of the High Treason Incident in May 1910 in which socialists and anarchists allegedly attempted to assassinate the emperor [6], these groups were further targetted. The Special Higher Police Division, known as Tokkō, established in Tokyo in August 1911 and subsequently in other prefectures, was created to monitor and control individuals and groups deemed to be a threat to the national polity and to conduct "thought control" through censorship. The subjects of their surveillance broadened to also include communists, liberals, labour activists, levellers movement activists, student movement activists, (core Japanese and colonial) nationalists, foreign residents and visitors (especially from the Soviet Union), returnees (especially from the Soviet Union) and religious groups [20].

In addition to Tokkō, the Imperial Japanese Army's military police (Kempei-tai) functioned not only to police military personnel but also acted as "security police" under the authority of Article 1 of Imperial Ordinance No. 337 (Kempei Ordinance)of October 1898. This allowed and authorised the Kempei-tai to monitor speech and behaviour of ordinary people and control their thought in order to keep the nation ready for war (in peace time) or to maintain the war-footing (in war time). Kempei-tai conducted some strict crackdowns against communists and socialists [11]. Both Tokkō and Kempei-tai police used preventive arrest and detainment of suspects as well as torture to elicit confessions.

From July 1928, prosecutors specialising in "thought crimes" (Shisō Kenji) were placed in each local prosecutor's office in the main cities following the roundup of around 1,600 members of the Japanese Communist Party on 15 March 1928. Shisō Kenji performed a complementary role to Tokkō in surveillance of security risks. In addition, they studied the theories of communism and socialism and developed effective techniques and probation systems to re-educate people into "true Japanese" attitudes [18].

These activities by the Tokkō, Kempei-tai and Shisō Kenji were authorised under a series of peace preservation laws such as the Public Security Preservation Law (Act No. 46 of 22 April 1925; Act No. 54 of 10 March 1941). These laws were presented to the populace as measures to prevent revolutionary threats to the established order while allowing universal male suffrage, by preventing the creation of revolutionary parties (particularly communist party) while initially preserving modest individual freedom of speech. However, the original law was soon found to be very limited in controlling political propaganda against the existing system. Consequently, the law was revised twice and the resultant "evil law" allowed the police agencies to monitor any member of the public and to support strongly the conduct of the Asia-Pacific War, even after Japan's continued losses [15].

2.2 After the War

The Human Rights Directive (Memorandum on Removal of Restrictions on Political, Civil, and Religious Liberties) issued by the General Headquarters of the Allied Forces (GHQ) on 4 October 1945 required (a) the repeal of laws which limited freedom of thought, religion, assembly and speech, (b) the dismantling of all organisations which had engaged in thought control including Tokkō, (c) the dismissal of the Home Minister, other top police commanders and all Tokkō police officers, and (d) the immediate release of political prisoners. Consequently, Tokkō was dismantled on 6 October, 353 political criminals and 1,896 people who were on probation for political crimes were freed by 15 October, and the Public Security Preservation Law and the

Public Order and Police Law were repealed on 15 October and 21 November, respectively. 4,990 people who were regarded as involved in Tokkō were dismissed. Kempei-tai was eliminated on 16 January 1946 [15, 19]. In reality a significant number of people who were involved in Tokkō escaped dismissal, and almost all of the Shisō Kenji continued their careers as prosecutors after the war, despite the abolition of the office by the Ministry of Justice on 15 October 1945 acting without GHQ orders [18].

The GHQ directive was understood as a measure to democratise and liberalise Japan and encouraged people who were suppressed during the war to claim their place in civil society. However, the resultant surge of social and labour movements and of the Japanese Communist Party were regarded as disruptive of public order by conservative Japanese politicians, and soon came to be viewed as dangerous by GHQ in the early stages of the Cold War. The Public Security Division (Kōan) was established in the Security Bureau of the Home Ministry on 19 December 1945 (and in each prefectural police department afterwards) once again to protect democracy (i.e. the status quo) from the threat of violent social and labour movements. In June 1946, Kōan started to collect domestic intelligence that they consider vital to national security and with GHQ's approval continually increased their personnel and intelligence-gathering capabilities. Despite the Japanese police structure being significantly changed by the dismantling of the Home Ministry in December 1947 and the full-scale revision of Police Act (Act No. 162 of 8 June 1954) in June 1954, Kōan's organisation and surveillance activities were consistently expanded and reinforced to control supposedly anti-democratic and anti-social activities until the early 1970s [20, 2].

It is alleged that Kōan inherited ideas, principles, technique, expertise and know-how concerning their intelligence activities from Tokkō, and in fact many of Tokkō police worked for Kōan after their purge was lifted [2]. Kōan is regarded as the most successful intelligence agency in Japan. Their ability and human resources are significantly greater than other similar agencies like the Public Security Intelligence Agency and the Cabinet Intelligence and Research Office [21, 17]. Kōan maintains a centralised system of command with the Security Bureau of the National Police Agency at the top, and monitors a broad range of individuals and groups deemed to be a security risk including the Japanese Communist Party, Communist Sympathisers (comsymps), trade unions, rightist groups, far-left militant groups, the General Association of Korean Residents in Japan, cult groups[1], radical environmental groups, anti-globalisation groups, anti-imperial system activists, espionage agents and Muslims [24]. In 2010 information was leaked from a division of the Tokyo Metropolitan Police (TMP) containing personal information on Muslims which had been gathered through various surveillance means. 17 people sued the police for invasion of privacy. In January 2015 the Tokyo District court awarded damages of over ¥90m to the plaintiffs (a decision upheld by the Higher Court in April 2015). However, the award was in respect of the leak and prompted by the TMP's poor data security and auditing. The court upheld the authority of the police to conduct the surveillance, despite the blanket nature of the surveillance which appeared to target any Muslim who came to the attention of the TMP for any reason [3]. This attitude by both police and courts emphasises the claim by Aoki [2] that Kōan engages in "general information gathering" monitoring ordinary people who are not deemed to be a security risk.

[1]including Aum Shinri-kyo (and its successor Aleph) which committed serious terrorist attacks using sarin gas [16]

The advancement of information and communication technology and the adoption of related laws such as Act on Wiretapping for Criminal Investigation (Act No. 137 of 18 August 1999) and Act on the Protection of Specially Designated Secrets (Act No. 108 of 13 December 2013) seem to have enhanced Kōan's capability of surveillance. As discussed in by the Asahi Shimbun in the report on the Muslim surveillance case [3], the recent adoption of the severe and wide-ranging new state secrecy law (Act on the Protection of Specially Designated Secrets) is an additional concern in this area, given that police surveillance activities are highly likely to be classified and are clearly within the scope of this act.

Furthermore, while seven war contingency laws were enacted in 2003 and 2004, the National Security Council was set up in 2013 and the re-interpretation of Japan's pacifist constitution to allow a "right of collective self-defence" is being pushed by Prime Minister Abe. The Japan Self-Defence Force (SDF) Intelligence Security Command (ISC) was established in August 2009 as a counterintelligence agency despite concerns that it could play a role similar to Kempei-tai. Its predecessor organisation, the Japan Ground Self-Defence Force (JGSDF) ISC undertook surveillance against peaceful anti-war activists protesting against SDF forces deployment to Iraq in 2004 [11].

The Japanese mass media rarely report on surveillance issues, such as Snowden's revelations, and as we shall see, awareness of and concern about these issues in Japan is much more limited than in other countries.

3. OVERVIEW OF THE SURVEYS

An initial pilot version of the survey was deployed in four Japanese universities and one Spanish university in June 2014. 491 responses were obtained in Japan and 50 from Spain. The survey results showed striking differences in the attitudes towards privacy, freedom, safety and security of the societies and individuals, state surveillance and Snowden's revelations between the two countries, although the sample size in Spain was limited. Follow-up interviews with 20 students at Meiji University conducted in July 2014 highlighted their confidence in government agencies and distrust in private companies in terms of privacy protection, underestimation of the threats of high-tech monitoring and unsympathetic attitudes towards Snowden regarding him as a reckless rebel against the state.

Based on the outcomes of those surveys, an online survey using a revised questionnaire was conducted in October and November 2014 among students at twenty-nine highly-rated (for teaching and/or research) Japanese universities. 1820 valid responses were received (out of 1887 submitted). All respondents held Japanese citizenship (two held dual citizenship with the US). The gender and age distribution of the respondents are shown in Table 1. 33.7% of respondents

(614/1820) majored in commerce/business administration, 18.3% (333/1820) in informatics, 10.0% (182/1820) in law, 8.3% (151/1820) in economics, 7.6% (139/1820) in sociology, 6.5% (119/1820) in policy making and 5.8% (106/1820) in technology/engineering.

Table 1: Respondent attributes (number (%))

Gender	**Male**				**Female**			
	1130 (62.1%)				690 (37.9%)			
Age	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**
	151 (8.3%)	436 (24.0%)	566 (31.1%)	355 (19.5%)	159 (8.7%)	72 (4.0%)	29 (1.6%)	52 (2.9%)

The questionnaire used in this survey consists of three parts plus an fact sheet about Snowden's revelations. The first part was answered by all of the respondents and included questions related to the right to privacy and privacy concerns. The second part asked respondents who had indicated that they already knew about Snowden's revelations about where they had obtained their information and whether they had discussed it with others, and whether they had changed their behaviour because of it. Even those who had said they knew about Snowden's revelations were then asked to read the author's brief (as neutral as they could write) description of Snowden's revelations. The third part then asked respondents about their attitudes to Snowden's actions, in particular seeking to replicate some of the questions asked by Pew Research of Americans [25].

After conducting the survey, the authors carried out follow-up semi-structured interviews with 56 of respondents at Meiji and Ehime Universities in May/June 2015.

4. SURVEY RESULTS AND DISCUSSION

4.1 Japanese Circumstances Related to Snowden's Revelations

4.1.1 Attitude towards the Right to Privacy in Japan

This section provides an overview of respondents' privacy attitudes, knowledge of and reactions to Snowden's revelations.

Respondents' privacy attitudes were studied by asking about their opinion on the perceived importance and understanding of the right to privacy. As shown in Table 2, on the one hand 93.7% of respondents (1539 of 1642) answered that the right to privacy was "very important" (37.4%) or "important" (56.3%), on the other hand, over half of respondents (58.4%; 942 of 1614) indicated that they "hardly" (56.9%) or "don't" (1.5%) understand that right. Moreover, the majority of the respondents who felt that the right to privacy was "important" (56.4%; 841 of 1491) were among those who indicated that they did not understand it, while more than nine out of ten respondents who did not understand the right (91.3%; 841 of 921) nevertheless felt that it is important (see Table 3). These results show that privacy is an emotional desire for Japanese young people rather than an intellectually understood legal right (actual or aspirational). Since 2008 the authors and other colleagues have carried out various surveys and interviews in Japan on the right to privacy and online privacy issues (using similar or identical questions). These results are consistent with the outcomes of those [22, 14, 23]. Many respondents in free text answers to surveys or in interviews admitted that although they often regard the right to privacy as important, they are only vaguely aware of what it entails. They report that their belief that it is important is a reaction to mass media reports and/or to high school or university ICT classes which stress the importance of privacy without providing a deep understanding of it.

Table 2: Frequency table of Q10 and Q13

Q10. Is your right to privacy important?		Q13. How well do you understand what the right to privacy is?	
Answer	Freq. (%)	Answer	Freq. (%)
Very important	614 (37.4%)	Understand very well	48 (3.0%)
Important	925 (56.3%)	Understand	624 (38.7%)
Not so important	96 (5.8%)	Hardly understand	918 (56.9%)
Not important at all	7 (0.4%)	Don't understand at all	24 (1.5%)
Total	1642	Total	1614

The survey results reveal Japanese youngsters' feeling of and attitude toward privacy invasion, via two approaches: perceived risk levels associated with activity, and perceived risk associated with organisations/technologies. Q6, Do you feel that your use of the Internet involves taking risks with your privacy?, had 20.3% of respondents (363 of 1791) answer "strongly" and 60.7% (1088 of 1791) answer "to an extent". So, over eighty percent of respondents felt their privacy to be under threat when using the Internet. Q7, Do you feel that your non-Internet activity involves taking risks with your privacy?, shows that only just over half of respondents perceived a privacy threat for their non-Internet activities (52.0%/931 of 1792). Figure 1 shows these difference as a comparative graph. If we take the four point scale as a quantitative evaluation by respondents of the level of risk from 0 ("Not at all") to 3 ("Strongly"), the mean score of Q6 (2.00) was higher than that of Q7 (1.56) and the difference (D=0.44) was significant at the one percent level ($t(1790)=26.712$, $p<.01$). In follow-up interviews respondents who had indicated that they felt their non-Internet activities were a risk to their privacy mentioned their own use of credit cards and of loyalty cards, and others' use of smartphone cameras, as worrying issues.

Table 3: Contingency table of Q10 and Q13

		Q13		
		Understand	Not Understand	Total
Q10	Important	650	841	1491
	Not important	19	80	99
	Total	669	921	1590

(The four-point scale answers to each question were transformed into two categories. In Q10, for example, "very important" and "important" were conflated to "important".)

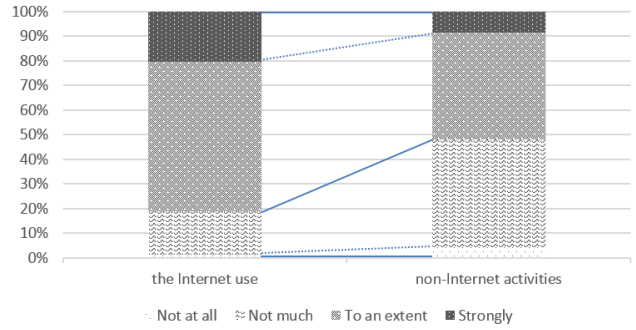


Figure 1: Risk recognition in Internet and non-Internet activities

In order to understand Japanese youngsters' perception of threats to their privacy, Qs 8 and 9 asked respondents to indicate the level of risk associated with different organisations and technologies, respectively. Again taking their responses (0="Not at all"–3="Very much") as quantitative Table 4 shows the average scores of each group as a source of privacy invasion. The top three privacy invasive groups were "Internet companies (2.17)", "telecom companies/Internet provider (1.78)" and "other for-profit companies (1.65)". The three types of government agencies had low averages: "law enforcement": 1.16, "secret service": 1.24 and "other government": 1.18, ranking 13th, 11th and 12th out of the 15 groups. Table 5 shows the average scores for the nineteen listed technologies as threats to privacy. "Smart phone (2.28)", "personal computer (2.20)" and "GPS (2.06)" ranked highest. Several online service technologies such as "social media services (1.94)", "online shopping (1.92)" and "online auction (1.92)" also had high average scores. The technologies with the lowest means were "home-based health monitoring (0.88; 19th)", "home automation witch senses human activities (0.89; 18th)" and "personal body monitoring (0.96; 17th)".

54 of the 56 respondents who took part in follow-up interviews had given survey responses indicating that smartphones and PCs were technologies that threatened their privacy. Of these, however, more than 70% (40 of 54) had not changed any of the privacy settings on the devices that they used. Nearly 60% (32 of 54) did not undertake any of the commonly recommended privacy-enhancing steps for PC or smartphone use such as deleting cookies, search and browser history or changing passwords. Over 75% (41 of 54) had not installed anti-virus software on their smartphones. On the other hand, nearly 70% (37 of 54) turned off or limited location-based services on their phones.

The survey results mentioned above indicates that Japanese youngsters tend to feel higher risk of privacy invasion for activities, organisations and technologies which have a direct association with the Internet use. Meanwhile, the threats posed by other types of organisation (e.g., non-profit organisations and government agencies) and technologies (e.g., devices for health management, motion sensors and video games) seem to be underestimated by Japanese respondents.

Very few follow-up interviewees had significant understand-

Table 4: Ranked means (0:low; 3: high) of 15 groups as perceived privacy threat

Q8. How much do you feel that the following groups threaten your privacy?		
Group	Mean	S.D.
Internet companies	2.17	.792
Telecom companies/ Internet providers	1.78	.839
Other for-profit companies	1.65	.880
Computer software companies	1.58	.855
Individuals who you don't know	1.53	.934
System Integrators	1.53	.875
Individuals who you know but not well	1.51	.783
Computer hardware companies	1.45	.849
Individuals who you know well	1.42	.881
Educational institutions	1.35	.884
Secret service government agencies	1.24	.919
Other government agencies	1.18	.879
Law enforcement government agencies	1.16	.886
Other not-for-profit organisations	1.16	.831
Health-care organisations	1.15	.861

Table 5: Ranked means (0:low; 3: high) of 19 technologies as perceived privacy threat

Q9. How much do you feel that the following technologies threaten your privacy?		
Technologies	Means	S.D.
Smart phone	2.28	.789
Personal computer	2.20	.807
GPS (Global Positioning System)	2.06	.848
Social media services	1.94	.927
Online auction	1.92	.951
Online shopping	1.92	.902
Making payments online	1.86	.932
Online games	1.67	.922
CCTV	1.62	.829
Smart card	1.41	.871
Behavioural targeting	1.37	.940
RFID (Radio Frequency Identification)	1.19	.846
Automatic Number Plate Recognition	1.12	.789
Portable video game console	1.07	.847
Smart meter	1.03	.758
Home video game console	1.02	.813
Personal body monitoring	0.96	.794
Home automation which senses human activities	0.89	.790
Home-based health monitoring	0.88	.770

ing of the organisation or activities of the three types of government agency involved in surveillance activities.Very few even knew of the existence of Kōan or the Investigative Department of the National Tax Agency. Several interviewees had the view that both the police and intelligence agencies in Japan were well-intentioned which acted to enhance societal security, and so they did not need to worry about their activities, since they would not do anything wrong. The majority of interviewees felt that hospitals, schools and other NPOs were trustworthy without needing or having any assurances about their intentions or operations.

4.1.2 The Degree of Recognition of and Interest in Snowden's Revelations in Japan

Japanese youngsters' recognition of and interest in Snowden's revelations were examined based on three points: recognition level of the revelations, information source for getting and/or updating knowledge of the revelations; and information relation activities.

Before the survey, 43.3% of Japanese respondents (680 of 1572) had heard about Snowden's revelations. This percentage was the lowest amongst our surveyed countries and much lower than most others (Germany: 98.6%, Sweden: 93.3%, China: 76.4%, New Zealand: 69.1% and Spain: 60.4%; Mexico: 46.7% and Taiwan: 46.5% were similar). Furthermore, the knowledge level of respondents who had heard the revelations was low. Only 27.3% respondents knew "a lot (2.8%)" or "a fair amount (24.5%)" about the contents of the revelations, 35.2% respondents knew "a lot (4.4%)" or "a fair amount (30.8%)" about the US government reactions to Snowden's revelations and 19.2% respondents knew "a lot (2.2%)" or "a fair amount (17.0%)" about the current status of Mr Snowden. This shows that Snowden's revelations are not well known among young people in Japan.

Mass media such as TV news (81.7%: 561 of 687), news on the Internet (44.3%: 304 of 687) and newspaper articles (34.5%: 237 of 687) were the main channels through which Japanese youngsters found out about Snowden's revelations, while personal communication channels were rarely the first contact with the information: social media: 10.9% (75 of 687); lectures at university: 7.0% (48 of 687); talks with friends/acquaintances: 2.8% (19 of 687).

Over eighty percent of respondents who knew about the revelations had not discussed it with their friends (82.7%: 563 of 681) and had not searched for more information about it (81.3%: 551 of 678). The fact that Japanese respondents mainly gathered information from mass media and a high percentage of them did not bother to gather additional information via active information search or having a discussion with others indicates that Japanese youngsters tend not to elaborate their knowledge and that Snowden's revelations did not consciously seem that relevant to their lives.

In follow-up interviews, those who had heard about Snowden's revelations reported an inability to understand or act upon them and/or that such things were irrelevant to their lives. The activities disclosed by Snowden seemed as though they came from another world or from a movie. They reported that they found it hard to imagine why Snowden had decided to act as he did. This was consistent with their general attitude to political and social issues: almost all of the interviewees reported a lack of interest in such things, and that it was not "cool" to discuss the Snowden revelations with their friends because it was such a non-issue.

4.1.3 Evaluation of Attitudes in Japan to Snowden's Activities

Thirty percent of respondents avoided judging the public value of Snowden's revelations, that is "no opinion" to they answered Q28 (Have Snowden's revelations served the public interest or harmed it?). But, of respondents who gave a judgement on the social contribution of the revelations, more than sixty percent admitted Snowden's activities had positive effects on the public interest (60.6%: 636 of 1049). Furthermore, over half of respondents felt that Japanese individuals should not need to give up privacy and freedom in order to ensure the safety and security of society and individuals (55.8%: 706 of 1265). However, they were not so optimistic about the actual impacts of the revelations. Only 17.8% of respondents answered Q36 (What social changes do you think have happened because of Snowden's revelations?) with "some social changes have happened" (258 of 1452). A large majority of of respondents (69.0%; 1002 of 1452) could not make a clear judgement (selected "no opinion") on the impact on society and 13.2% respondents (192 of 1452) reported "no" impact. These results seem to indicate that whereas the majority of Japanese youngsters have a positive opinion of Snowden's actions, they tend not to feel that they have had a strong impact on their society.

42 of the follow-up interviewees were willing to express an opinion about surveillance generally in Japan. Of these, 32 believed that "suspicious characters" (which to them included cultists, ex-convicts, gangsters and all foreigners in Japan) should be monitored by government agencies. They believed that they themselves would never become targets of state monitoring. A few were of the opinion that Muslims living in Japan should be kept under 24/7 surveillance.

4.2 Empirical Consideration about Influence of Snowden's Revelations

4.2.1 Do Snowden's Revelations Have Any Influence over Risk Perception of Privacy Invasion?

Two research questions are considered concerning the impact of Snowden's revelations on young people's attitudes and actions on privacy and surveillance: RQ1) Did Snowden's revelations have any influence over risk perception of privacy invasion?; RQ2) Was there any difference in actions taken to protect privacy between those who understood well or poorly about Snowden's revelations?

First, we divided the respondents into two groups in terms of whether they had heard about Snowden's revelations ("Heard" Group) or not ("Not Heard" Group) using the response to Q19 (Have you heard about Snowden's revelations?). Two other survey questions were considered via T-test with "Not Heard" as "control" and "Heard" as "treatment" groups.

RQ1: Did respondents who had heard about Snowden's revelations tend to recognise more risk of privacy invasion compared to those who did not know the revelations?

The average score of answers to Q6 (Do you feel that your use of the Internet involves taking risks with your privacy?) from the "Heard" group (2.06, SE=.026) exceeded that of "Not Heard" group (1.95, SE=.021), with a statistically signicant difference at the one percent level (D=.11, 95% CI [.042, .170]; t (1407.717) = 3.322, p<.01). Those who had heard about Snowdens revelations reported feeling more at risk of privacy invasion in their online activity than the group who had not heard.

RQ2: Did respondents who had heard about Snowden's revelations regard government agencies as greater privacy threats than those who had not heard?

The average scores of Qs 18-m (How much do you feel that law enforcement government agencies threaten your privacy?) and Q18-n (How much do you feel that secret service government agencies threaten your privacy?) and Q18-o (How much do you feel that other government agencies threaten your privacy?) were compared using a T-test with respect to the "Heard" and "Not Heard" groups (with answers treated as numeric (0="not al all"–3="very much")

The T-test for all three questions about the perceived privacy risk from government agencies shows that the "Heard" group were on average more concerned than those in the "Not Heard" group, all at a one percent significance level:

Law Enforcement Government Agencies
Heard: M=1.23, SE=.038; Not Heard: M=1.10, SE=.031; D=.131, 95% CI [.038, .230]; t (1205.34) = 2.656; p<.01
Secret Service Government Agencies
Heard: M=1.31, SE=.039; Not Heard: M=1.17, SE=.033; D=.140, 95% CI [.040, .243]; t (1214.851) = 2.748; p<.01
Other Government Agencies
Heard: M=1.26, SE=.037; Not Heard: M=1.13, SE=.031; D=.130, 95% CI [.027, .227]; t (1213.451) = 2.651; p<.01

The results of these data analyses indicate that Snowden's revelations have had significant influence over the perception of privacy invasion in Japan, even though in general the Japanese respondents tended not to regard government agencies as a serious threat to their privacy.

4.2.2 *Differences in Actions in Response to Snowden's Revelations Dependent on Reported Level of Understanding*

Converting the four options for answers to Q23 (How much do you know about the contents of Snowden's revelations?) into binary categorical data (i.e. respondents who answer "A lot" or "A fair amount" are categorised as "High-knowledge group" and those who answer "Not much" or "Little" are labelled as 'Low-knowledge group"), the relationship between respondents' knowledge level about Snowden's revelations and actions in response to the revelations (Q24) was considered via a Chi-square test. The result shows that these two variables are independent (Chi-square(1) =0.508, p>0.1; ϕ=-0.028, p>0.1), so there appears to be no evidence of impact on the actions of Japanese respondents depending on their self-evaluation of their level of knowledge about Snowden's revelations.

Table 6: % "yes" to Qs30/33 in eight country cases

	JPN	CHN	DEU	ESP	MXC	NZL	SWE	TWN
Q30	13.9	39.2	41.3	59.7	63.2	54.2	47.7	50.9
Q33	15.3	25.8	59.6	62.9	64.2	62.5	67.2	56.4

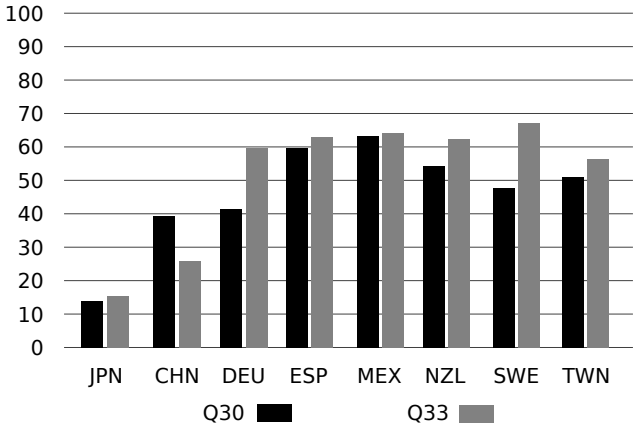


Figure 2: % "yes" to Qs30/33 in eight country cases

Many of the follow-up interviewees expressed a doubt that they needed to change their practices in response to Snowden's revelations, because they were not doing anything wrong. This fits with the earlier reported attitude that respondents felt both trusting of the good will of at least Japanese agencies which might be monitoring them, and that they had little understanding of the issues surrounding foreign monitoring by the NSA/GCHQ.

4.2.3 *Would Japanese Young People Follow Snowden's Lead?*

Although purely hypothetical questions are of course difficult for respondents to answer and when faced with the reality, many might choose differently, questions 30 (If you were an American citizen and were faced with a similar situation to Snowden, do you think you would do what he did?) and 33 (If you were faced with a similar situation to Snowden in Japan, i.e. you found out that a Japanese intelligence agency was conducting similar operations to those of the NSA and GCHQ, would you, as a Japanese citizen or a resident in Japan, do what he did?) help us to build a view of the attitudes to state surveillance amongst young people. In particular they provide an interesting point of comparison between countries as to how strongly young people feel about such government activities. In Japan, only 13.9% (169 of 1212) (Q30: US situation) and 15.3% (174 of 1134) (Q33: Japanese situation) of respondents reported that they believed they would take the same actions as Snowden. As shown in Table 6 and Figure 2, Japanese young people are the low outliers amongst those studied. This is consistent with our other results that Snowden's revelations have had very limited conscious impact on Japanese young people.

5. CONCLUSION

The coverage of the Snowden revelations in Japanese mass media has been limited, shallow and often misleading. The presence in Japan of US military bases almost certainly car-

rying out NSA interception operations has not been mentioned, contrasting with mass media coverage in other countries where in addition to the NSA/GCHQ international surveillance questions, local issues such as cracking of local telcos (e.g. Belgium [4]) or surveillance of local politicians (e.g. Germany [5]) is given significant coverage.

Japanese young people are far less likely than their counterparts in other countries to have heard about Snowden's revelation, to know much about them if they have heard, or to have taken conscious actions in response. Despite this, those who have heard the revelations do exhibit significant though modest increases in concern about privacy.

The inclusion by follow-up interviewees of foreign residents of Japan as suitable subjects of surveillance and their placement alongside cultists, ex-convicts and gangsters, shows a strong streak of racism and xenophobia among Japanese university students. The statement by a few that Muslims should, simply by the nature of their religion, be under 24/7 surveillance shows that whatever the legality, ethicality or utility of the Tokyo Metropolitan police surveillance of Muslims [24], that the court ruling that this is valid is probably a reflection of public opinion (young people tend to be more liberal than older people and more highly educated people tend to be more liberal than less highly educated people, suggesting that such attitudes among university students are probably more prevalent in Japanese society in general).

Japanese young people are strong outliers internationally in their unwillingness to follow Snowden's example if placed in a hypothetical similar situation. As one respondent put it in a free text answer: "I don't stick my neck out for anyone".

6. ACKNOWLEDGEMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006 and the JSPS Grant-in-Aids for Scientific Research (B) 24330127 and (B) 25285124. 65 academics from Universities around Japan helped in encouraging their students to respond to our survey. There is no space to list them here, but the authors extend their sincere thanks for those efforts.

7. REFERENCES

[1] A. A. Adams, K. Murata, and Y. Orito. The Japanese Sense of Information Privacy. *AI & Society*, 24(4):327–341, 2009.

[2] O. Aoki. *Japanese Security Police*. Kodansha, Tokyo, 2000. In Japanese.

[3] Asahi Shimbun. Court: Police can gather personal information on Muslims. tinyurl.com/pcdmp6n, 2014.

[4] G. Burton. GCHQ hacked Belgian telco Belgacom. tinyurl.com/p7x2m6w, 2013.

[5] Der Spiegel. Embassys Espionage: The NSA's Secret Spy Hub in Berlin. tinyurl.com/q5xafmm, 2013.

[6] S. Giffard. The development of democracy in Japan. *Asian Affairs*, 27(3):275–284, 1996.

[7] G. Greenwald. *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. Metropolitan Books, 2014.

[8] M. Gurnow. *The Edward Snowden affair: Exposing the politics and media behind the NSA scandal*. Blue River Press, Indianapolis, IN, 2014.

[9] L. Harding. *The Snowden files: The inside story of the world's most wanted man*. Vintage Books, New York, NY, 2014.

[10] H. Ito. *Commentaries on the Constitution of the Empire of Japan*. Chūō Daigaku, 2nd edition, 1906. Translated by M. Ito.

[11] A. Koketsu. *Military Police Politics: The Age of Surveillance and Intimidation*. Shinnihon Shuppansha, Tokyo, 2008. In Japanese.

[12] D. Lyon, editor. *Theorizing Surveillance: The Panopticon and Beyond*. Willan, Cullompton, 2006.

[13] E. Moglen. Privacy under attack: The NSA files revealed new threats to democracy. tinyurl.com/kcocf7o, 2014. 27th May.

[14] K. Murata, Y. Orito, and Y. Fukuta. Social attitudes of young people in japan towards online privacy. *Journal of Law, Information and Science*, 23(1):137–157, 2014.

[15] S. Nakazawa. *The Public Security Preservation Law: Why Did Party Politics Create the Evil Law?* Chuokoron-Shinsha, Tokyo, 2012. In Japanese.

[16] National Police Agency (Japan). 1996 White Paper on Policing: Addressing new types of organised crimes: A review of the Aum Shinrikyo cases. tinyurl.com/qzlj37w, 1996. In Japanese.

[17] H. Noda. *The Deep Structure of the Public Security Intelligence Agency*. Chikumashobo, Tokyo, 2005. In Japanese.

[18] F. Ogino. *Thought Prosecutors*. Iwanami Shoten, Tokyo, 2011. In Japanese.

[19] F. Ogino. *The Special Higher Police*. Iwanami Shoten, Tokyo, 2012. In Japanese.

[20] T. Ogura. *Electronic Government and Surveillance-Oriented Society*, chapter 13, pages 270–295. In Lyon [12], 2006.

[21] Y. Omori. *Japanese Intelligence Agencies*. Bungeishunju, Tokyo, 2005. In Japanese.

[22] Y. Orito, Y. Fukuta, and K. Murata. I will continue to use this nonetheless: Social media srvive users' privacy concerns. *International Journal of Virtual Worlds and Human Computer Interaction*, 2:92–107, 2014.

[23] Y. Orito, K. Murata, and Y. Fukuta. Do online privacy policies and seals affect corporate trustworthiness and reputation? *International Review of Information Ethics*, 19:52–65, 2013.

[24] M. Oshima. *Who is Kōan Monitoring?* Shinchosa, Tokyo, 2011. In Japanese.

[25] Pew Research Center. Obama's NSA Speech Has Little Impact on Skeptical Public. tinyurl.com/nw2fpfs, 2014.

[26] Pew Research Center. Public Perceptions of Privacy and Security in the Post-Snowden Era. tinyurl.com/p2536wh, 2014.

[27] K. Shoji and M. Sugai. The Ashio copper mine pollution case: the origins of environmental destruction. In Ui [28], chapter 1, pages 18–63.

[28] I. Ui, editor. *Industrial Pollution in Japan*. United Nations University Press, Tokyo, 1992.

Young People Do Care — Snowden's Revelations Have Had an Effect in New Zealand

Gehan Gunasekara
University of Auckland
Private Bag 92019, Victoria Street West
Auckland 1142, New Zealand
+64 9 9235218
g.gunasekara@auckland.ac.nz

Kiyoshi Murata
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2165
kmurata@meiji.ac.jp

Andrew A. Adams
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2329
aaa@meiji.ac.jp

Ana María Lara Palma
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
+34 947 259 360
amlara@ubu.es

ABSTRACT
This study investigates the attitudes towards and social impacts of Edward Snowden's revelations in New Zealand through a questionnaire survey and follow-up interviews with New Zealand youngsters as part of the worldwide cross-cultural analyses. The survey results showed striking contrasts with those in other countries reflecting New Zealand's socio-cultural and political environment.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues – *abuse and crime involving computers, privacy, use/abuse of power*

General Terms
Security, Human Factors, Legal Aspects

Keywords
Edward Snowden, privacy, state surveillance, social impact, New Zealand

1. INTRODUCTION
The disclosures, starting on 5th June 2013, made by former NSA contractor Edward Snowden revealed effectively limitless information gathering and indiscriminate mass monitoring carried out by the NSA (National Security Agency), an intelligence agency of the US Department of Defence, and its British counterpart the GCHQ (Government Communications Headquarters) [1]. Amongst the revelations were that the agencies had directly accessed the servers of United States companies, including Microsoft, Google, Yahoo!, Facebook, Apple, YouTube and Skype [1]. The remarkable exposes resounded throughout the world as countless millions of individuals use these companies.

Although there was, on the one hand, outrage regarding Snowden's disclosures by some sectors of society, others have instead echoed the statement of Sun Microsystems' Scott McNealy that "You have zero privacy anyhow. Get over it" [2, 3].

In New Zealand, the Snowden case was given much publicity but was not the only one causing disquiet to those concerned about privacy. Domestic incidents of illegal spying by intelligence agencies also came to light in 2013. In addition, the Government adopted several new laws and amended others greatly extending the power of intelligence gathering in the course of that year. Despite privacy again being an election issue for all the opposition parties, the Government was re-elected in 2014 with an increased majority. Although enquiries have been initiated as a result of the ongoing Snowden revelations concerning New Zealand's participation in intelligence gathering, the extent of popular concern after 2013 remains in doubt [4].

The more pertinent question for researchers is the extent to which individuals have modified their behaviour as a consequence of the Snowden affair as well as the extent to which his conduct was seen as justified. Tentative research is mixed as to the first question, but there are indications from the United States that behaviour modification has indeed occurred [5]. Privacy scholars are well-accustomed to the "chilling" effect of surveillance [6].

A related question is awareness, particularly amongst the young, of the facts and motivations surrounding Snowden's conduct, as well as its ethical implications. Snowden refers, in interviews, to

Conference'10, Month 1–2, 2010, City, State, Country.

his motivations in part being the "Internet Principle of the seven headed hydra", in other words the expectation that other whistle-blowers will be emboldened by his conduct [7]. Such a desire may, however, be fanciful given the personal repercussions for Snowden of his conduct [1, 7]. This study, although small in comparison to its international counterparts, attempts to evaluate these questions by sampling the views of undergraduate New Zealand University students as to the implications for them of the Snowden affair and its ethical perspective.

2. STATE SURVEILLANCE IN NEW ZEALAND

New Zealand is the world's youngest society as it was first settled by Polynesian peoples (Maori) only in the fourteenth century [8]. Its relatively short history, however, has seen remarkable developments not paralleled elsewhere. In the nineteenth century, conflict between indigenous Maori and European settlers over land and fierce resistance by Maori resulted in both political compromise and grievances that resonate to the present day. New Zealand became arguably the world's first true democracy as both women and indigenous people were given full electoral rights well before the end of the nineteenth century [8]. In addition, the country became known for social innovation, such as the provision of old age pensions and welfare well before other western nations. Some of these innovations, such as the country's universal no-fault accident insurance scheme have, in turn, created vulnerabilities for individuals as vast amounts of personal data is collected and processed by public sector agencies [9].

The Cold War saw concerns about subversion in New Zealand leading to the setting up of the country's domestic intelligence agency, the New Zealand Security Intelligence Service (NZSIS) in 1956. This agency exists to the present day although its track record is somewhat mixed, its occasional excesses and blunders attracting both criticism and satire [10]. The country's agency charged with gathering foreign intelligence and with the interception of foreign communications, in addition to safeguarding the New Zealand Government's own communication, the Government Communications Security Bureau (GCSB) remained largely in the shadows until its powers and oversight mechanisms were laid out in the Government Communications Security Bureau Act 2003. Despite New Zealand's exclusion from the ANZUS military alliance with the United States after enactment of the New Zealand Nuclear Free Zone, Disarmament, and Arms Control Act 1987, intelligence co-operation has continued with New Zealand being an active participant in the "Five Eyes" network [11].

The 2013 Snowden disclosures were joined in New Zealand by the revelation that the GCSB had exceeded its powers by intercepting domestic communications of New Zealand residents. This was brought to light during court proceedings involving Internet tycoon and entrepreneur Kim Dotcom, a New Zealand resident, whom the US Federal Bureau of Investigation (FBI) was seeking to extradite on criminal copyright piracy charges. Furthermore, the New Zealand Government introduced changes to the Act governing the GCSB extending its powers to conduct domestic surveillance particularly in relation to metadata. The Telecommunications (Interception Capability and Security) Act 2013 also obliged network providers to provide back doors in their systems for intelligence agencies to be able to use their interception powers. The changes were implemented despite significant public protest and criticism from intellectuals [12].

Notably, New Zealand is one of only three countries that lack a written constitution [13]. A constitutional monarchy, its constitutional rules are found in a mix of ordinary statutes, precedents found in court decision as well as unwritten conventions or practices that are unenforceable but are invariably followed in practice. Despite concerns raised at the potential for abuses of executive power [14], New Zealand has only experienced occasional periods where civil liberties have been restricted, notably during periods of large-scale industrial unrest in the early and mid-twentieth century [8]. The introduction of proportional representation, modelled on Germany, in 1996 has also acted as a significant check on executive power as no single party is generally able to obtain a majority in the legislature, thus forcing compromise and empowering smaller parties [15].

A further important milestone is the New Zealand Bill of Rights Act 1990 (NZBORA). Despite not containing a specific right to privacy, this charter of fundamental freedoms contains a suite of rights such as the right not to be subjected to unreasonable search and seizure and the right to due process. The NZBORA does not have the status of higher law and cannot be used to challenge other legislation but has still been used to restrain surveillance by the State.

For example, Kim Dotcom argued the authorities breached his rights under the NZBORA [16]. In 2007 New Zealand experienced its first domestic terrorism case with the prosecution of a group including Maori separatists accused of planning to conduct terrorist acts in New Zealand. Evidence had, however, been obtained illegally by Police by video surveillance obtained by trespassing on private land. The evidence was found by the Supreme Court to be inadmissible and charges brought under the Suppression of Terrorism Act 2002 against the defendants were dropped as a consequence [17, 18]. Although temporary legislation and, ultimately, the Search and Surveillance Act 2012 legitimised the types of surveillance that had occurred, these were not backdated due to public outcry [19].

Both the 2007 terrorism case and the Kim Dotcom case attracted widespread media coverage in New Zealand and, together with the Snowden disclosures, created a climate of interest in privacy throughout 2013. Opinion polls consistently showed high degree of anxiety concerning privacy in the online environment but also a greater trust in government handling of personal data than reposed in the handling of such data by private corporations [20]. However such polls have not specifically addressed the question of mass surveillance, the ethical aspects of the Snowden affair and whether any behaviour modification has resulted from the disclosures.

3. SURVEY RESULTS AND DISCUSSION

3.1 Overview of Survey

The survey conducted in New Zealand (at the University of Auckland) elicited 66 responses. As is the case with the surveys done in Japan, China, Taiwan, Mexico, Spain, Germany and Sweden, 39 questions were formulated to test the effect of Snowden's disclosures on New Zealand youngsters as well as their attitudes towards the disclosures. The responses to the questions are discussed under the following headings: the nature of the responders, the attitudes towards the right to privacy demonstrated, the degree of knowledge and interest in Snowden's revelations, evaluation of Snowden's conduct and, last but not least, whether the revelations have had any influence over perceptions of risk concerning privacy invasions.

3.2 Nature of Responders

The response group was 39% male and 61% female. Their mean age was 20 with responders tending to be younger with only 15% being over 25 years old. As the survey was only conducted at the University of Auckland, New Zealand's largest University, it accounted for 100% of responders. As the survey was primarily aimed at business students 75% of responders to the survey were studying business/commerce while the next largest group, technology/engineering accounted for only 7%.

New Zealand's population is ethnically diverse with significant Asian (mainly Chinese) and Polynesian minorities, although these are considerably over-represented in the Auckland metropolitan area. Of the respondents, 55% indicated New Zealand as their nationality as opposed to 45% who did not. The survey used the categories found in census data to identify the ethnic groups respondents identified with, with the ability to identify more than one category [21]. The largest single group was Chinese with 30% with New Zealand Europeans accounting for only 20%. However, the category "other" accounted for 35% being the second largest group.

3.3 Attitudes towards the Right to privacy in New Zealand

In response to whether their use of the Internet involved taking risks with privacy 53% felt that it did to an extent and 31% felt strongly that it did. Only 17% felt that Internet use involved not much or no risk (Figure 1). By contrast, 60% of respondents felt non-Internet activity involved little or no risks to privacy, whilst 41% indicated that it did to an extent or strongly (Figure 2).

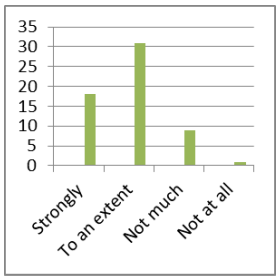


Figure 1. Internet Threats

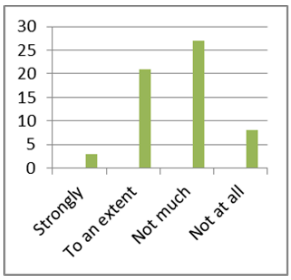


Figure 2. Non-Internet Threats

As to which groups threatened privacy most, the data collected from respondents showed that much higher levels of concern were expressed about threats from Internet companies, secret service agencies, telecommunications and Internet providers, as opposed to much lower levels of concerns regarding individuals known to the respondents and not for profit organisations. The outcomes are shown in Table 1 each respondent being asked to rank the threats with 3 being highest and 0 lowest (excluding no opinion/response options).

Table 1. How Much Groups Threatened Privacy (3: high; 0: low)

Q8. How much do you feel that the following groups threaten your privacy?		
Group	**Mean**	**S.D.**
Internet companies	2.42	0.88
Secret service government agencies	2.02	0.98
Telecom companies/ Internet providers	1.94	1.08
Computer software companies	1.69	1.27
Law enforcement government agencies	1.69	1.24
Other for-profit companies	1.64	1.52
Other government agencies	1.61	1.33
System Integrators	1.54	1.52
Health-care organisations	1.50	1.30
Educational institutions	1.47	1.61
Individuals who you don't know	1.45	1.37
Individuals who you know but not well	1.43	1.49
Computer hardware companies	1.41	1.44
Other not-for-profit organisations	1.35	1.63
Individuals who you know well	1.33	1.47

Questions concerning the technologies that threatened privacy the most yielded interesting results. Most threatening were smart phones, social media, personal computers and GPS technologies, whereas portable video game consoles, home video consoles and home automation technologies featured as less threatening. These categories are depicted in the following table.

Table 2. Technologies Respondents Believed Threatened Privacy (3: high; 0: low)

Q9. How much do you feel that the following technologies threaten your privacy?		
Technologies	**Mean**	**S.D.**
Smart phone	2.31	1.02
Social media services	2.23	0.88
Personal computer	2.17	1.00
GPS (Global Positioning System)	2.02	1.04
CCTV	1.89	1.14
Making payments online	1.89	1.20
Online shopping	1.82	1.21
ANPR (Automatic Number Plate Recognition)	1.69	1.21
Behavioural targeting	1.63	1.26
Online auction	1.41	1.41
Smart meter	1.35	1.47
Online games	1.34	1.47
Smart card	1.21	1.57
Personal body monitoring	1.20	1.63
RFID (Radio Frequency Identification)	1.19	1.63
Home-based health monitoring	1.15	1.75
Home automation which senses human activities	1.13	1.70
Home video game console	1.07	1.73
Portable video game console	0.83	2.03

On the all-important question as to whether the right to privacy is important an overwhelming 96% answered that it was important or very important. As to whether respondents understand what the right to privacy is, on the other hand, 58% did claim to understand whereas 42% claimed they did not. More important, however, was

analysis of the open-ended questions of both groups as to what they felt privacy meant and why it was important.

From these responses, several clear themes could be identified. The predominant one was the importance of control over the sharing of personal information with 35%. This was followed very closely by concerns about personal safety or security with 33%, many respondents stating concern as to identity theft and the potential for harm should personal data be misappropriated. The third largest group of responses, accounting for 16%, identified privacy as important as it was an essential human right, the word "democracy" being used by some. This was closely followed by responses grouped under the ability to make real choices with 14%. These used terms such as the individual's right to be "independent", "freedom", "individuality" and the "risks" that might attach from the choices made by an individual. Only one response stated that privacy was important in maintaining trust and relationships with those with whom information was shared. Finally, only one respondent answered why privacy was *not* important by referring to the trade-off between privacy and connectivity. The qualitative responses are depicted in the following diagram.

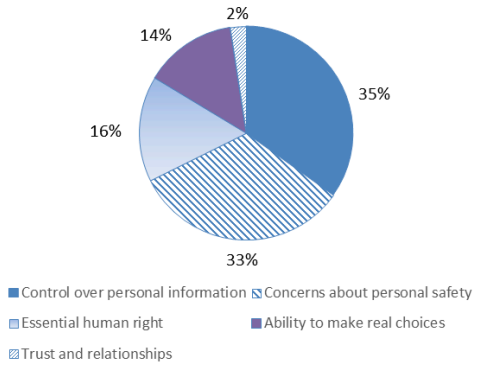


Figure 3. Why Privacy Is Important

Although there were fewer responses describing what the right to privacy is – around half those for why it was important – similar themes are evident in both sets of responses. The vast majority, 55% referred to the control of information related to themselves. Some 23% referred to it being a democratic right or human right whilst 18% referred to it giving individuals the right to do what they wanted provided they were acting lawfully. The free will to make decisions concerning themselves and the right to be "left alone to decide" matters were responses falling within this description. Only one response identified the protection of sensitive aspects of an individual's life such as entry into their home or their voting records. These qualitative responses are shown below.

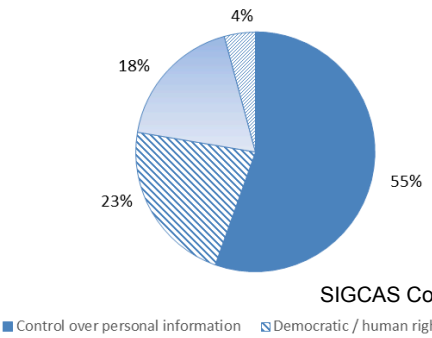


Figure 4. What Is the Right to Privacy

Despite the small size of the sample, the themes that emerge from these open-ended questions are extremely significant. In the first place, the predominance of information or data privacy shows that awareness of its significance in the digital age is high amongst youngsters. In follow-up interviews respondents stated in some cases they had learnt from personal experience the harm that can result from misappropriation of personal data. Others expressed fear as to the consequences – such as career-wise – of information about them being misconstrued or used against them.

The ability to make real, as opposed to manipulated, choices also figured throughout with terms such as being "independent", "freedom", "individuality" and the ability to "take risks" being values to which respondents subscribed. This included being "left alone to decide" and being "free to make decisions" concerning oneself. In this respect the responses are in line with the views of scholars who have pointed to an important function of privacy being protecting autonomous lives and individual autonomy [22, 23].

As far as respondents felt that about New Zealand individuals having to give up privacy and freedom in order to ensure safety and security of the society and individuals the opinions were nuanced, with a significant majority of the view that this was necessary to an extent with a normal distribution as to views but with a tendency to value privacy when faced with having to give it up very much as opposed to some extent. This can be seen in the following graph.

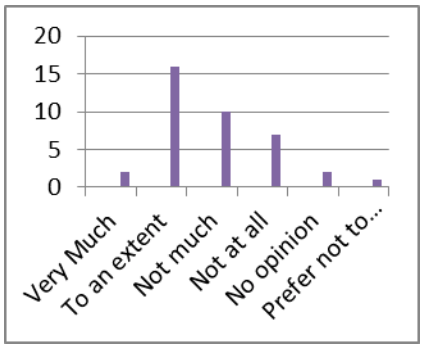


Figure 5. Privacy versus Security

The final question assessing attitudes to privacy concerned the degree of knowledge respondents had about a number of both domestic and foreign intelligence and law enforcement agencies as well as agencies devoted to protecting human rights and privacy. The results revealed that there was better knowledge of foreign agencies generally than domestic ones. Likewise, there was more knowledge of intelligence and law enforcement agencies (such as the FBI and NSA) than of agencies dedicated to the protection of human rights such as the Office of the Privacy Commissioner (Table 3).

Of those who had heard of the agencies the question further evaluated the extent of understanding of the agencies. The results were revealing as, again, greater knowledge appeared to exist of

the foreign agencies than of domestic ones. Of the respondents, on average 53% understood the foreign agencies as opposed to 29% for the domestic ones.

Table 3. Awareness of Agencies

Agency	Have not heard of it		Have heard of it	
FBI	1	2%	22	98%
CIA	3	7%	20	93%
NSA	6	14%	19	86%
GCHQ	24	57%	4	43%
Government Communications Security Bureau	16	38%	10	62%
Office of the Ombudsman	24	57%	5	43%
Office of the Privacy Commissioner	18	43%	6	57%

Although no statistically significant difference was found between the ethnic groups identified in the survey as to attitudes to privacy, significant differences were evident in the degree of recognition and understanding shown by the different ethnic groups as to the agencies listed. These differences are depicted in the following diagram.

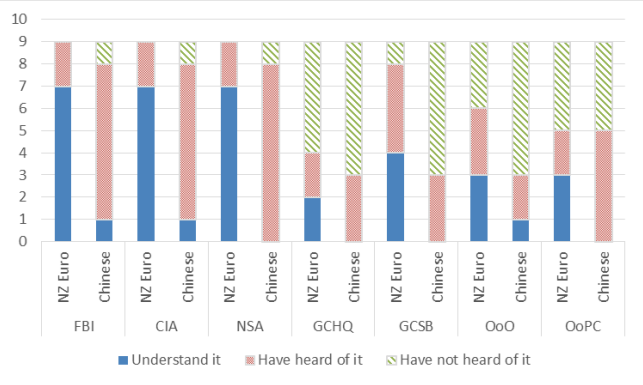


Figure 6. Understanding and Knowledge of Organizations

3.4 Knowledge and Interest in Snowden's Disclosures

The questions went on to assess the extent of knowledge as to Snowden and Manning disclosures and respondent's evaluation of their effects and consequences for the whistle-blowers. As far as the earlier whistle-blower US Army Private First Class Bradley (Chelsea) Manning is concerned 55% had heard of this, but only 30% claimed to know a lot or fair amount concerning it. By comparison, 70% had heard about Edward Snowden's revelations although a similarly low 34% claimed knowledge about their contents.

The questions further asked how respondents obtained and updated their knowledge surrounding Snowden's revelations. The highest sources were listed as Internet news reports, television news reports followed by social media. Interestingly, the least frequent source was university lectures. On the other hand, whilst 59% of respondents had talked about the affair with others, only 39% had searched for further information about Snowden's revelations. As to knowledge as to the current status of Snowden himself, only a minority were well informed with 37% knowing a lot or fair amount as opposed to 62% knowing not much or little. Likewise 45% knew a lot or fair amount about the US Government's reactions to the revelations, whereas 55% knew either not much or a little.

The open-ended questions as to respondents' opinions as to Snowden's motives yielded a range of views with a few pointing to self-interest or financial motives but the vast majority identifying a number of more public-spirited reasons. The single most prominent of the latter was openness or transparency: the secrecy of the NSA programmes being seen as their most pernicious aspect as they led to a complete lack of accountability for the agencies concerned. Related motives were the misuse of the anti-terror campaign to undertake mass surveillance. Finally, some responses identified "justice' and "public duty" and "public interest" as motivational factors.

3.5 Evaluation of Snowden's Conduct

The final series of questions addressed attitudes towards the ethicality of Snowden's disclosures. The vast majority, 79% stated they had served the public interest whereas 16% preferred not to express an opinion at all. Only 5% felt they had harmed the public interest. A logical corollary was to ask whether the US Government ought to pursue a criminal case against Snowden. In this regard 48% of respondents choose not to answer but of those who did 15% said yes whilst 85% said no.

The questions then progressed to a somewhat more personal level by asking whether, if the respondents were American citizens faced with a similar situation to Snowden, they would do what he did. Interestingly, the largest number 37% chose not to answer this question but of those who did 54% said they would and 46% indicated they would not. The questions further asked whether had they found out that a New Zealand intelligence agency was conducting similar operations to those of the NSA and GCHQ, they would, as New Zealand citizens, do what he did. The same percentage chose not to answer this question but of those who did, a larger proportion, 62.5% said they would act in the same way.

The difference between the responses to these parallel questions can be explained when the open-ended answers are evaluated, as respondents clearly articulated a more benign view of New Zealand's judicial and political system than that of the United States. To begin with few respondents, 20%, answered the question why they would do what Snowden did if they were US citizens with only a slightly higher number answering why they would do so if they were New Zealand citizens. Themes evident in the responses included ethical/moral ones and particularly the need to expose secrecy and lack of accountability. One poignant response, which could have been made by Snowden himself, was that Government was supposed to be for the people but mass surveillance of Americans proves the interests of Government are served instead.

Reasons for not acting in the same manner as Snowden predictably emphasised the risk to personal wellbeing ("I would rather not be exiled to Russia") through having to spend life on the run. A few, however, emphasised Snowden's own lack of trust and misuse of classified information entrusted to him. More revealing were the reasons given by those who would act as Snowden did if the actions had been in New Zealand.

These revealed a far more sanguine view of New Zealand society and attitudes towards whistle-blowers than that which is perceived to be the case in the United States. A common view was the perceived lack of corruption in New Zealand's judicial system and the actively participatory nature of its democracy: it was felt to be a national duty to expose deception on the part of elected officials. The "less serious consequences" view is reflected in one response as follows: "New Zealand is an incredibly free country so the risk of getting arrested and harshly punished would be significantly less than what would happen if I did so in the USA." Such views may be indicative of New Zealand's unique cultural and social environment. The qualitative responses are depicted in the following diagram.

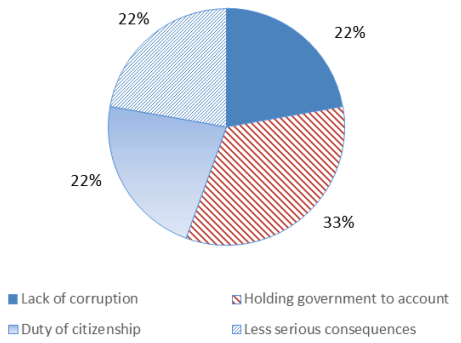


Figure 7. Reasons for Whistleblowing in New Zealand

3.6 Social Effects of Snowden's Revelations

The final two questions addressed respondents' views as to the social effects of the Snowden disclosures and asked whether they had modified their own behaviour as a consequence. It first asked what social changes they thought had occurred because of the revelations. Forty per cent chose not to respond but of those who did 43% said there had been no change while 57% said there had been with these being listed in their open-ended answers. In order of frequency of responses (greatest to least) social changes identified by respondents as a consequence of Snowden's actions were: greater caution by individuals in their online actions, greater awareness and discourse surrounding privacy issues, the cynical view that effects are only short term, politicians making empty promises and, finally, distrust of government.

Secondly the questions asked if respondents had changed their way of communicating online using systems such as social media (such as Twitter, Facebook, Messenger, YouTube, blogging, Skype, email and instant messaging) since they had heard about Snowden's revelations. More than one answer was allowed. Interestingly, all but one respondent answered this question, unlike many of the other questions, and while 41% said they had not changed their behaviour at all 59% said they had altered their conduct in a variety of ways. These responses are shown in Figure 8.

Only 6% of respondents, however, elaborated further as to the changes they had undertaken. These were: deciding to never start a social media account, use of a VPN, deleting a significant portion of Facebook posts, pictures, friends, and all personal information including last name and birthday and seeking privacy awareness software and services and use of encryption. Follow-up interviews also highlighted a small number of individuals who stated they had removed historic Facebook posts and were more savvy on information exchanged with app providers, but due to the small size of the interview sample (10% of survey respondents) it is difficult to state whether such conduct is symptomatic of more general behaviour.

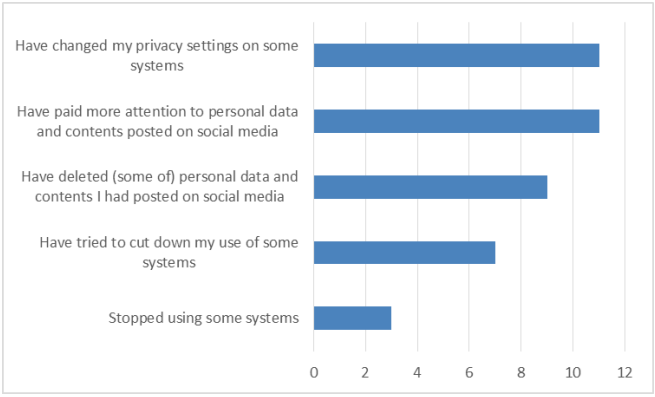


Figure 8. Changes in Behaviour

4. CONCLUSIONS

This tentative study of the effects of the Snowden affair on young New Zealanders has yielded some interesting conclusions. In the first place, respondents viewed online behaviour as carrying greater privacy risks than off line activities. Secondly, Internet, for profit companies and intelligence agencies were seen as posing a greater threat than not for profit organisations and known individuals. Thirdly, respondents were divided as to whether privacy needed to be given up to ensure security with the majority prepared to give it up to an extent but few prepared to do so to a great extent. Fourthly, there was greater awareness of overseas intelligence and law enforcement agencies than New Zealand ones, and poor knowledge in particular of human rights agencies in New Zealand such as the Office of the Privacy Commissioner and the Ombudsman.

The most important data, however, relate to the opinions of respondents as to the value of privacy and their understanding as to what it is. There was little difference between the ethnic groups who were surveyed in the sample as to their support for the concept of privacy which was overwhelmingly endorsed. In addition respondents articulated surprisingly lucid explanations as to why privacy was important and what the right to privacy is. Respondents were clearly cognizant of the crucial role personal data plays in the digital age. Notably, privacy was seen as an aspect of democracy, freedom and personal autonomy, thus placing New Zealand youngsters' views within mainstream privacy thinking. This may be somewhat reassuring from the standpoint of older generations accustomed to these concepts.

Finally, whilst respondents had generally heard of the Snowden disclosures far fewer claimed to have more specific knowledge as to the contents of the disclosures. There was overwhelming support for his actions and opposition to his having to face criminal prosecution. Whilst a majority of respondents stated they would emulate his actions, the proportion who would do so had the actions been against New Zealand intelligence agencies was higher with respondents having greater confidence in the treatment of whistle-blowers in New Zealand. These results assist

in providing an answer to the question, referred to earlier, posed by Snowden as to the motives for his actions [7].

5. REFERENCES

[1] Greenwald, G. 2014. *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. Metropolitan Books, New York, NY.

[2] Schwartz, J. 2001. As Big PC Brother Watches, Users Encounter Frustration. *New York Times* (Sep. 2001), C6.

[3] The end of privacy. *Science*, Special Issue (Jan. 2015). Available at: http://www.sciencemag.org/site/special/privacy/index.xhtml (Accessed on May 22 2015).

[4] Kirk, S. 2015. GCSB will be investigated over claims New Zealanders spied on in Pacific. *Stuff.co.nz* (Mar.2015). Available at: http://www.stuff.co.nz/national/politics/67518446/gcsb-will-be-investigated-over-claims-new-zealanders-spied-on-in-pacific (Accessed on May 22 2015).

[5] Rainie, L. and Madeen, M. 2015. Americans' Privacy Strategies Post-Snowden (Mar. 2015). Available at: http://www.pewinternet.org/2015/03/16/Americans-Privacy-Strategies-Post-Snowden/ (Accessed on May 22 2015).

[6] Solove, D. 2008. *Understanding Privacy*. Harvard University Press, Cambridge, MA.

[7] Poitras, L. 2014. Citizenfour. Praxis Films. Aailable at: http://www.praxisfilms.org/films/citizenfour (Accessed on June 26 2015).

[8] King, M. 2003. *The Penguin History of New Zealand*. Penguin, Auckland, New Zealand.

[9] KPMG 2012. *Independent Review of ACC's Privacy and Security of Information* (Aug. 2012). Available at: https://privacy.org.nz/assets/Files/Media-Releases/22-August-2012-ACC-Independent-Review-FINAL-REPORT.pdf (Accessed on May 22 2015).

[10] Rudman, B. 2015. The GC(SB): A touching story of everyday spies. *The New Zealand Herald* (May 2015). Available at: http://www.nzherald.co.nz/opinion/news/article.cfm?c_id=466&objectid=11451304 (Accessed on May 22 2015).

[11] Hager, N. 1996. *Secret power*. Craig Potton Publishing, Nelson, New Zealand.

[12] New Zealand Law Society 2013. GCSB Bill remains flawed despite proposed changes (Aug. 2013). Available at: https://www.lawsociety.org.nz/news-and-communications/news/august-2013/gcsb-bill-remains-flawed-despite-proposed-changes (Accessed on May 22 2015).

[13] The United Kingdom, Israel and New Zealand.

[14] Palmer, G. 1979. *Unbridled power? : An Interpretation of New Zealand's Constitution and Government*. Oxford University Press, Wellington, New Zealand.

[15] Palmer, G. 1997. *Bridled Power: New Zealand Government under MMP*. Oxford University Press, Auckland, New Zealand.

[16] *Kim Dotcom and others v United States of America* [2014] NZSC 24.

[17] Radio New Zealand News 2013. IPCA criticises illegal searches during Urewera raids. *Radio New Zealand News* (May 2013). Available at: http://www.radionz.co.nz/news/national/135737/ipca-criticises-illegal-searches-during-urewera-raids (Accessed on May 22 2015).

[18] *Hamed v R* [2011] NZSC 101, [2012] 2 NZLR 305.

[19] Video Camera Surveillance (Temporary Measures) Act 2011.

[20] Office of the Privacy Commissioner 2014. *Individual privacy & personal information UMR Omnibus Results* (Mar. 2014). Available at: www.privacy.org.nz/news-and-publications/surveys (Accessed on 22 May 2015).

[21] 2013 Census. Available at: http://www.stats.govt.nz/Census/2013-census.aspx?gclid=CLXf96D008UCFUoJvAodHaoABQ (Accessed on 22 May 2015).

[22] Rossler, B. 2005. *The Value of Privacy*. Polity, Oxford, UK.

[23] Westin, A. 1967. *Privacy and Freedom*. Atheneum, New York, NY.

The View from the Gallery: International Comparison of Attitudes to Snowden's Revelations about the NSA/GCHQ

Andrew A. Adams
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
aaa@meiji.ac.jp

Kiyoshi Murata
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
kmurata@meiji.ac.jp

Yasunori Fukuta
School of Commerce
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
yasufkt@meiji.ac.jp

Yohko Orito
Faculty of Law and Letters
Ehime University
3 Bunkyo-cho, Matsuyama
Ehime 790-8577, Japan
orito.yohko.mm@ehime-u.ac.jp

Ana María Lara Palma
Civil Engineering Department
Management Area
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
amlara@ubu.es

ABSTRACT

The series of revelations made by Edward Snowden revelations starting on 5th June 2013 exposed a true picture of state surveillance or, more precisely, surveillance conducted by an industrial-government complex in the democratic nations. His revelations have attracted heavy doses of both praise and censure; whereas some have positively evaluated his deed as an act of valour to protect democracy against the tyranny of the state, others have criticised him as a traitor to his country that have been preoccupied with responses to the threat of terrorism since the 9.11 attacks. Indeed, the US government filed charges of spying against him on 21st June, and he is forced to live in exile in Moscow. He said that only the American people could decide whether sacrificing his life was worth it by their response [10]. The Pew Research Foundation found in a survey that although Americans are deeply split on whether Snowden's actions served or harmed the public interest, that younger groups regarded his actions as more beneficial than harmful when compared with older groups

Inspired by the Pew Research Foundation's surveys [13, 14], an international group of academics led by the authors of this paper have conducted surveys on young people (students at their universities) about their attitudes to privacy online, and the actions of Bradley/Chelsea Manning and Edward Snowden in separate and different modes of grand leaks. This survey has been deployed in China, Germany, Japan, Mexico, New Zealand, Spain, Sweden and Taiwan. with further deployments expected.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues: abuse and crime involving computers, privacy, use/abuse of power

General Terms

Security, Human Factors, Legal Aspects

Keywords

Edward Snowden, privacy, state surveillance, social impact

1. INTRODUCTION

Edward Snowden's revelations about the operations of the US' NSA (and its primary partner the UK'S GCHQ, as well as the intelligence agencies in Canada, Australia and New Zealand) which started on 5th June 2013 confirmed many of the worst fears of privacy/anti-surveillance activists and academics, and even some of those previously dismissed as conspiracy theory nonsense. Both his act of revelation and the activities he exposed have attracted heavy doses of both praise and censure; whereas some have positively evaluated his deed as an act of valour to protect democracy against the tyranny of the state, others have criticised him as a traitor to a country that have been preoccupied with responses to the threat of terrorism since the 9.11 attacks. Indeed, the US government filed charges of spying against him on 21st June 2013, and he has been living in exile in Moscow. He said that only the American people could decide whether sacrificing his lifestyle was worth it by their response [10]. However, it is clear that the issue Snowden raised is not just for American citizens.

Lively discussions of national security, safety and security of societies, personal freedom and privacy have been gener-

ated in many countries by Snowden's revelations with many books and papers recently published [10, 5, 6, 7, 13, 14]. However, there seem to be differences in press and government reactions, so the question of ordinary people's attitudes towards the revelations, and, therefore, their social impacts in different social contexts [1].

Inspired by the Pew Research surveys of Americans' attitudes to Snowden, a group of international academics led by the authors and including colleagues from Germany (DEU), Mexico (MXC), Spain (ESP), Sweden (SWE) and New Zealand (NZL) (see section 6 for details) carried out similar surveys in their own countries (Germany, Japan (JPN), Mexico, Spain, Sweden and New Zealand). The Japanese team (with help from students studying in Japan) also conducted surveys in Taiwan (TWN) and the People's Republic of China (the PRC (CHN)).[1] Due to resource limitations, unlike the original Pew Research Center survey, only university students were recruited as respondents.

This paper presents first a process analysis of the development of such a survey, in particular the challenges of localising the survey to different countries. In addition, the initial comparative analysis of the results from some of these countries is presented. Other papers at the conference present more detailed analyses of individual country results.

1.1 Social Background of the Study Countries

Most of the countries so far studied [2] can be reasonably regarded as having had an authoritarian government within living memory. Sweden and New Zealand have long democratic histories, albeit a colonial one in New Zealand. Japan and Germany have been regarded as democratic since the 50s. Taiwan, a colony of Japan from the late nineteenth century until the end of the second world war, was then subject to the military rule of the mainland China-exiled Kuomintang until 1987, with the first presidential election only happening in 1996. The PRC remains a one-party state. Spain was a military dictatorship from 1939 to 1975 and a transition to a democratic government in 1981. Mexico was a one-party state from 1929 until the mid-80s. A gradual introduction of multi-party elections from the 70s through to the 90s finally led to the election of a president from an opposition party in 2000. Most respondents (most being under 25) in most countries except the PRC, therefore, have not had direct personal experience is life under an authoritarian regime, although for many the residual effects of such regimes may well be significant.

2. DEVELOPING AN INTERNATIONAL COMPARATIVE SURVEY

As noted above, the inspiration for this international set of surveys was a survey of US' citizens attitudes to Snowden's actions [3, 13, 14]. That survey covered adult citizens of all ages from young to retirement. It also considered US political viewpoints amongst the participant attributes gathered and used to inform the analysis. Primarily for resource reasons, as this work was conducted as a small part of modestly funded projects, the initial decision was taken to focus efforts on gathering responses from university students. This is a well-known drawback of much academic social science research, using students as proxies for young people in general and for people of all ages. The results of these studies should therefore be treated carefully with respect to their broader applicability to the general populations (non-university attending young people and older adults). The second purely practical concern was for the countries chosen to be studied, which was limited by both linguistic issues and availability of academic partners willing to deploy the survey amongst their students.

Some of the surveys were translated into the relevant local language. The New Zealand, Swedish, Spanish and Mexican versions were deployed in English. The German survey was translated into German. Broadly similar translations into Mandarin Chinese were used for the PRC and Taiwan, though using simplified and traditional Chinese hanzi characters respectively. The Japanese survey was translated into Japanese.

The goal of the set of surveys was to provide a basis for international comparison of attitudes. Hence, the nationality of respondents was asked. In addition, to allow for the possibility of significant regional variations in attitudes to be identified in some cases, further ethnicity detail were sought either by free text box ("please specific your 'other' nationality") and/or a list of likely possible origins (for New Zealand, for example, eight ethnic identities were listed, with a free text box for "other" also available). This diversity (or lack thereof) in the student body was also reflected in the way questions were asked about respondents' willingness to emulate Snowden's actions. For all countries, respondents' were asked whether they believed they would emulate Snowden's actions if they were a US citizen. They were then also asked to consider the same hypothetical questions regarding their likely actions with respect to their own country and its intelligence agencies. For countries with very few overseas students, this was limited to a questions regarding that country (e.g New Zealand or Japan). For countries such as the UK with a significant number of foreign students this question had to be altered to separate UK and overseas citizens in the answer to the question about whether they would emulate Snowden regarding their own country.

3. INTERNATIONAL COMPARISONS

3.1 General Privacy Attitudes

In order to evaluate the relative attitudes to privacy between respondents in different countries, a one-way inter-country analysis of variance (ANOVA) was performed. The results of this are shown in Figure 1.

Numeric values were assigned to the textual answers to the question "Is your right to privacy important?" (answers: "not important at all"; "not so important"; "important"; "very important" allocated a linear numeric interpretation 0–3). The mean values of respondents' answers in each country were compared using a Welch test which indicated that there were at least some statistically significant differences at the one percent level.[3] Post-hoc multiple pair-wise com-

[1]The three letter form given in brackets is used in all tables and figures for that country.

[2]It is planned that the survey will also be deployed in the UK and Canada.

[3]Adjusted $F(7, 339.736) = 58.777$, $p < .01$

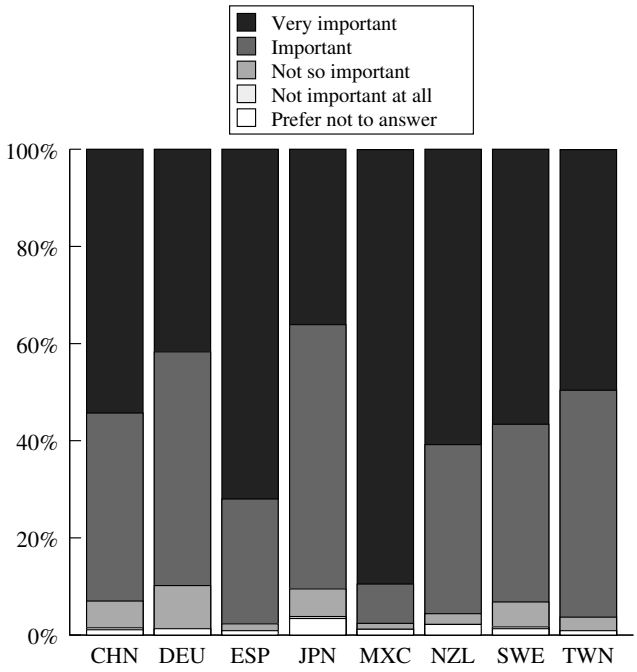


Figure 1: Is Your Right to Privacy Important?

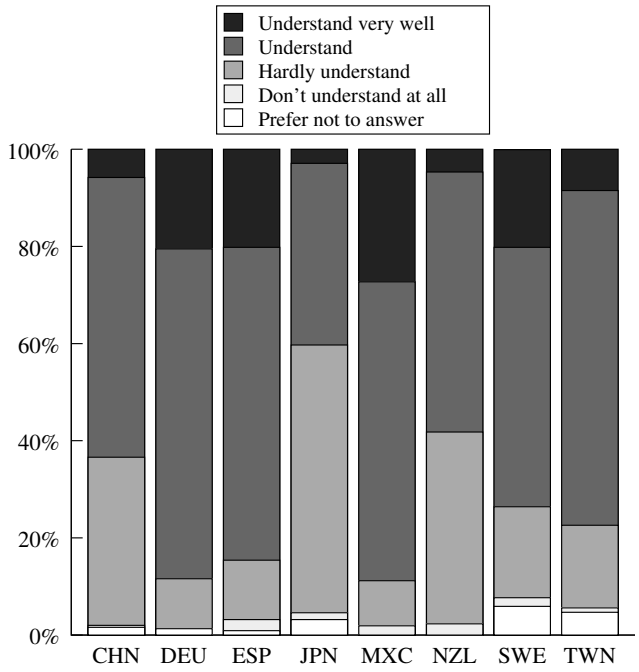


Figure 2: How Well Do You Understand What the Right to Privacy Is??

parisons were then undertaken to identify the pairs which had significant differences. As the answers within each country were not generally homogenous, Tamhane's T2 test was adopted instead of Tukey's test which requires homogeneity. The pairwise comparisons are shown in Table 1.

In addition to asking respondents for their evaluation of the importance of the right to privacy, they were asked to report their own perception of their understanding of that right("Don't understand at all"; Hardly Understand; Understand; Understand very well). The results of their answers are shown in Figure 2.

The same pairwise analysis was applied to this question to identify which countries had statistically significant differences regarding respondents' understanding of the right to privacy, again by assigning numeric values to the answers, computing a mean value and comparing means. The results are shown in Table 2.

Since the Snowden revelations were related to government surveillance, the survey asked respondents to evaluate the threats to their privacy posed by for-profit, not-for-profit and governmental organisations. Comparing the for-profit and government sector responses using paired samples t-tests, it was found that respondents from the PRC, Japan and Taiwan, were statistically more concerned about for-profit than government invasion of privacy at $p < .01$. In the other countries there was no statistically significant difference in concerns between the two sectors. See Table 3 for details.

3.2 Knowledge of Snowden's Revelations

The survey first asked respondents if they had heard about Snowden's revelations or not. If they indicated that they had, they were asked to evaluate their level of knowledge: "little"; "not much"; "a fair amount" or "a lot". Tables 4 and 5 shows the original percentages and numbers for the separate questions. Figure 3 shows the spread of answers as a percentage of respondents, interpreting the original answer as "nothing" and recalculating the percentages for the second questions as appropriate.

Table 4: Had They Heard about Snowden's Revelations?

		CHN	DEU	ESP	JPN	MXC	NZL	SWE	TWN
yes	%	76.4	98.6	60.5	43.3	46.7	69.0	93.3	46.5
	No.	188	72	130	680	70	29	194	47
no	%	23.6	1.4	39.5	56.7	53.3	31.0	6.7	53.5
	No.	58	1	85	892	80	13	14	54

Table 5: How Much Have You Heard about Snowden's Revelations?

		CHN	DEU	ESP	JPN	MXC	NZL	SWE	TWN
1	%	2.7	37.5	16.5	23.5	26.1	27.6	8.5	0.0
	No.	5	27	21	160	18	8	16	0
2	%	44.1	54.2	40.2	49.3	47.8	37.9	41.8	51.1
	No.	82	39	51	336	33	11	79	24
3	%	52.2	8.3	39.4	24.5	23.2	34.5	41.8	46.8
	No.	97	6	50	167	16	10	79	22
4	%	1.1	0.0	3.9	2.8	2.9	0.0	7.9	2.1
	No.	2	0	5	19	2	0	15	1

1 — little; 2 — not much; 3 — a fair amount; 4 — a lot

As can be seen from these various presentations, there is wide variation between countries on the self-reported knowl-

Table 1: Pairwise Country Comparison of Importance of Privacy (Difference of Means)

	CHN	DEU	ESP	JPN	MXC	NZL	SWE
DEU	.136						
ESP	-.233**	-.369**					
JPN	.176**	.040	.409**				
MXC	-.411**	-.546**	-.177**	-.587**			
NZL	-.089	-.225	.144	-.265	.322*		
SWE	-.042	-.178	.191*	-.218**	.369**	.047	
TWN	-.003	-.138	.230*	-.179	.408**	.086	.039

**) significant difference at $p < .01$ *) significant difference at $p < .05$
Positive: top row country had a higher mean; negative: the left column country had the higher mean

Table 2: Pairwise Country Comparison of Understanding of Privacy (Difference of Means)

	CHN	DEU	ESP	JPN	MXC	NZL	SWE
DEU	-.394**						
ESP	-.342**	.052					
JPN	.262**	.656**	.604**				
MXC	-.464**	-.070	-.122	-.726**			
NZL	.081	.474**	.422**	-.182	.544**		
SWE	-.276**	.118	.066	-.539**	.188	-.357*	
TWN	-.191	.202	.150	-.454**	.272**	-.272	.085

**) significant difference at $p < .01$ *) significant difference at $p < .05$
Positive: top row country had a higher mean; negative: the left column country had the higher mean

Table 3: Pairwise t-tests For-Profit (FP) and Government (G) Mean Privacy Concern

		CHN	DEU	ESP	JPN	MXC	NZL	SWE	TWN
FP	M	2.10	1.97	2.01	1.71	1.62	1.77	1.71	1.98
	SE	.037	.066	.053	.017	.054	.103	.042	.048
G	M	1.35	1.88	1.93	1.18	1.71	1.76	1.64	1.72
	SE	.050	.093	.073	.021	.072	.134	.055	.071
Stats	D	.746	.089	.083	.527	-.085	.015	.072	.261
	95% CI	[.638,.850]	[-.063,.240]	[-.039,.226]	[.490,.567]	[-.217,.045]	[-.218,.242]	[-.014, .152]	[.123,.400]
	t	(273) 13.34	(77) 1.17	(136) 1.26	(1595) 27.31	(154) -1.20	(48) .13	(254) 1.64	(101) 3.68
	p	$< .01$	$> .1$	$> .1$	$< .01$	$> .1$	$> .1$	$> .1$	$< .01$

edge of Snowden's revelations amongst respondents. Germany and Sweden are outliers in having very few who had not heard about them at all (DEU: 1/73(1.4%); SWE 14/208 (6.7%)). Nowhere did more than 10% of the respondents report knowing "a lot" about the revelations, with Sweden having the largest such group (15/204 (7.4%))[4].

[4]It should be noted that not all respondents who indicated that they had heard about Snowden's revelations answered the question regarding their amount of knowledge, for example 194 Swedish respondents replied that they had heard of the revelations but only 189 gave an evaluation of their level of knowledge.

3.3 Evaluation of Snowden's Actions

In their 2014 survey of US citizens' attitudes to the Snowden revelations the Pew Research Center reported that most young Americans regarded Snowden as having served the public interest [3]: "57% of 18- to 29-year olds said the leaks have served rather than harmed the public interest. . . ". Figure 4 shows the similar evaluation (at a more fine-grained response level) for these international surveys.

The evaluation of Snowden's actions was allocated to a four point scale (-2="harmed it a lot"; -1="harmed it to some extent"; +1="served it to some extent"; +2="served it a lot"). These analyses considered only those who expressed an opinion (the survey also gave respondents three options for not expressing an opinion: "no opinion"; "prefer not to answer"

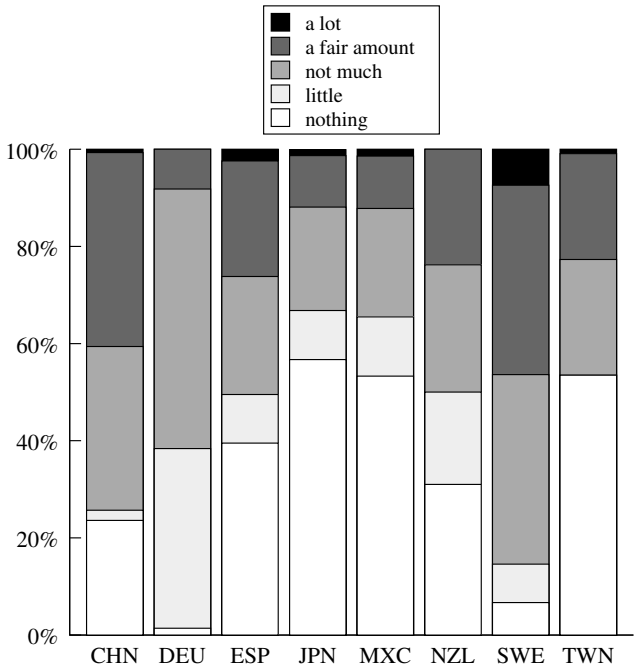


Figure 3: How much do they know about Snowden's revelations?

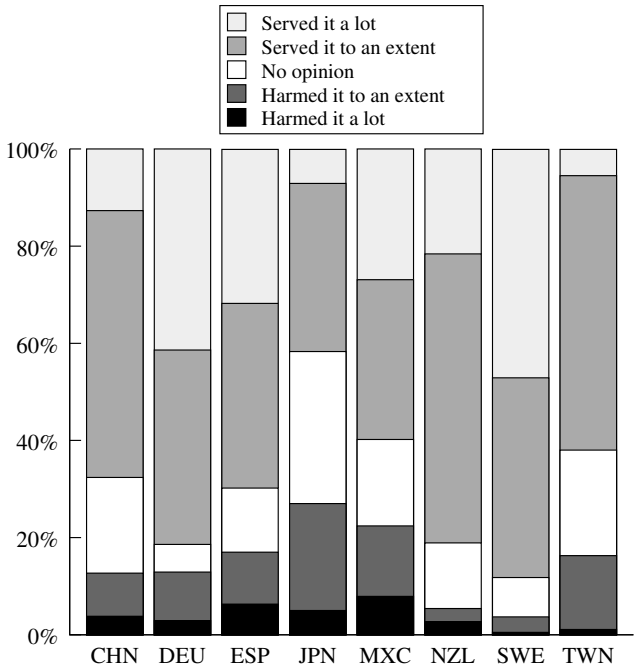


Figure 4: Did Snowden Serve or Harm the Public Good?

and not selecting any answer). The mean evaluation, given in Table 6, for all countries in these surveys was positive, though quite variable (as can also be seen from figure 4).

As Levene's test for homogeneity of variances indicated that the variance within each country was not demonstrably homogenous (Leven statistic (7, 1855) = 37.281, $p < .01$), a

Table 6: Mean scores for "Did Snowden Serve or Harm the Public Good?"

	CHN	DEU	ESP	JPN	MXC	NZL	SWE	TWN
Mean	0.80	1.14	0.90	0.24	0.69	1.09	1.42	0.64
SE	0.08	0.13	0.09	0.04	0.12	0.15	0.06	0.11

Welch test was applied to these means to check for the existence of statistically significant differences. This showed that there is at least one pairwise comparison with a difference significant at the one percent level.[5] Post-hoc pairwise comparisons, again using Tamhane's T2 test, give the results shown in Table 7.

Sweden is clearly an outlier in this, having a higher mean evaluation score for Snowden's actions that the PRC, Spain, Japan, Mexico and Taiwan, all at a one percent significance level. Japan is the outlier in the other direction, having a lower mean than all the other countries, at a five percent significance level compared to Mexico and Taiwan and a one percent significance level for all the others. This fits with an intuitive reading of Figure 4.

3.4 The Impact of Snowden's Revelations

In terms of their reactions to Snowden's revelations, a majority of respondents in all countries except Japan and Taiwan reported that they have changed their communication practices after hearing about Snowden (among those who had heard about them). Even in Taiwan, which had a fairly small sample size, approximately half (23 out of 47) of the respondents who had heard about Snowden's revelations had changed their practices. In Japan, by contrast, only a quarter (26.39%; 181 of 686) who had heard about Snowden's revelations reported that they had consciously changed their communication practices. A Chi-squared test confirms that Japan differs from all the other countries at a p<.01 level on this point (see Table 8). These results from Japan are more in line with those of the Pew Research Center regarding Americans' reactions to Snowden. In that survey of a broadly representative group of Americans only 34% reported changing their online communications behaviour in response to the Snowden revelations [15].

In Japan, there was a difference (significant at the one percent level) between the perceived privacy risk of Internet activity and from government law enforcement agencies and secret services (see [11] for the detailed statistical analysis of this). In contrast a similar analysis shows that in no other country[6] did knowledge or not of Snowden's revelations change those related privacy concerns at a statistically significant level.

3.5 Willingness to Emulate Snowden's Actions

Although purely hypothetical questions are of course difficult for respondents to answer and when faced with the reality, many might choose differently, the following questions help us to build a view of the attitudes to state surveillance

[5] Adjusted F(7,268.612)= 48.305, $p < .01$

[6] The results from Germany could not be subjected to a suitable test for significance due to the tiny proportion (1 of 73) who had not heard about Snowden's revelations.

Table 7: Pairwise Country Comparison of Evaluation of Snowden's Actions (Difference of Means)

	CHN	DEU	ESP	JPN	MXC	NZL	SWE
DEU	-.341						
ESP	-.104	.237					
JPN	.554**	.895**	.658**				
MXC	.107	.448	.211	-.447*			
NZL	-.298	.043	-.195	-.853**	-.406		
SWE	-.628**	-.287	-.525**	-1.182**	-.736**	-.330	
TWN	.156	.497	.260	-.398*	.049	.455	.785**

**) significant difference at $p < .01$ *)significant difference at $p < .05$
Positive: top row country had a higher mean; negative: the left column country had the higher mean

Table 8: Contingency table and Result of Chi-square test about Reactions in Web Communication

	Changed?		Statistics		
Country	Have changed	Haven't changed	$\chi^2(1)$	P-value	ϕ
JPN	181	505			
CHN	114	73	78.528	<.01	-.300 (p<.01)
DEU	45	27	40.615	<.01	-.231 (p<.01)
ESP	78	50	59.365	<.01	-.270 (p<.01)
MXC	52	18	68.355	<.01	-.301 (p<.01)
NZL	17	12	14.439	<.01	-.142 (p<.01)
SWE	116	76	77.617	<.01	-.297 (p<.01)
TWN	23	24	11.137	<.01	-.123 (p<.01)

amongst young people: "If you were an American citizen and were faced with a similar situation to Snowden, do you think you would do what he did?" (QUS) and "If you were faced with a similar situation to Snowden in your home country, i.e. you found out that your own government's intelligence agency was conducting similar operations to those of the NSA and GCHQ, would you, as a citizen or a do what he did?" (QL) . In particular they provide an interesting point of comparison between countries as to how strongly young people feel about such government activities. These results are shown in Table 9 and Figure 5.

Table 9: "yes"% to QUS/QL in Eight Country Cases

	CHN	DEU	ESP	JPN	MXC	NZL	SWE	TWN
QUS	39.2	41.3	59.7	13.9	63.2	54.2	47.7	50.9
QL	25.8	59.6	62.9	15.3	64.2	62.5	67.2	56.4

For the US hypothetical, a significant majority (at the one percent level) of Mexican respondents and (at the five percent level) of Spanish respondents indicated they would emulate Snowden. Among Chinese and Japanese respondents a significant majority (at the one percent level) indicated that they would *not* emulate Snowden. In all other countries there was no statistically significant majority either way.

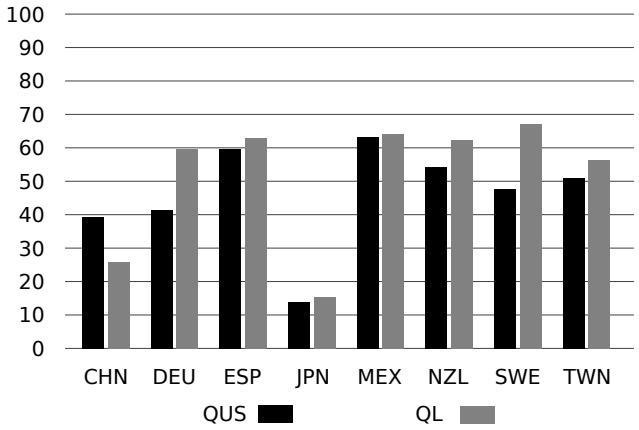


Figure 5: "yes"% to QUS/QL in Eight Country Cases

These results were mirrored in the local hypothetical variants (for this one the Spanish tendency to emulate Snowden had a higher one percent significant level as well) except that in the case of Sweden there was also a statistically significant majority (at the one percent level) in favour of emulating Snowden if faced with a Swedish equivalent scenario.

For each country, the answers to these two questions were checked for consistency using a Chi-square test. In most countries there was no statistical difference between the answers to the two questions. However, in Sweden and the PRC there was a difference significant at the one percent level, in opposite directions. In Sweden, respondents were more likely to emulate Snowden in the hypothetical Swedish case (Chi-square test for independence QUS/QL: Chi-square (1) = 10.281, $p < .01$). In the PRC, respondents were less likely to emulate Snowden in the hypothetical Chinese case (Chi-square (1) = 7.314, $p < .01$).

In addition to the relationships between willingness to emulate Snowden in the US or local situation, the relationship

between respondent's evaluation of Snowden's actions and their willingness to emulate him in the hypothetical US or local situations was also evaluated. The numeric score assigned to evaluations of Snowden's actions given in section 3.3 was again used to provide quantitative analysis.

In all countries, the mean evaluation of Snowden's actions among those who indicated that they would emulate Snowden was more positive than among those who indicated that they would not (in both the US and local hypothetical versions). However, in only a few cases was the difference statistically significant (according to a t-test using the answer to the evaluation of Snowden's actions as the test variable and the answer to the emulation question as the grouping).For the US hypothetical Spanish respondents had a significant correlation between their evaluation of Snowden's actions and their willingness to emulate him ($p = .05$). Japanese, Mexican and Swedish respondents were correlated at the one percent significance level ($p < .01$). For the local hypothetical Spain again showed a five percent significance of correlation ($p < .05$), while Japan, Mexico and Sweden had a correlation significant at the one percent level ($p < .01$). The sample sizes and distributions made tests on the German, Taiwanese and New Zealander respondents uncertain.

4. CONCLUSIONS

Government are fairly well trusted and statistically significantly more trusted[7] than the private sector in respect of privacy in the three SE Asian countries (Japan, the PRC and Taiwan), while in all the others there was no significant difference. This is an interesting result given that the PRC is an authoritarian regime, Taiwan has only recently moved into a relatively democratic system and Japan has been relatively democratic for over half a century.

Information and understanding about Snowden's revelations varies considerably with respondents in the PRC, Germany and Sweden indicating a higher level of knowledge than elsewhere. In the PRC, where much of the news is state controlled or heavily state influenced, it seems likely that the government there sees the revelations about US surveillance of its own (and others') citizens as a useful normalising factor for its own online surveillance and censorship regime. The revelations that the mobile phone of the German Chancellor (head of government) had been under surveillance by the US, an allied country, is one of the reasons why the Snowden revelations have received so much press coverage there. Both the former Nazi and East German (GDR) histories of heavy use of surveillance to oppress the population have also led to strong distrust of government surveillance systems [4]. Despite its democratic history Sweden has a history of public scepticism towards government dataveillance [4, 9], while the high profile criminal court case against Julian Assange, head of the Wikileaks organisation which published the confidential US government material released by Chelsea Manning, has probably increased press interest in the similar issue of Snowden's revelations. It appears that the lack and poor quality of information about Snowden available in Japan makes people unconsciously more worried about their privacy but unable to do anything about those concerns.

[7]See Table 3 for the details — note that this Table reports levels of concern, so a lower mean indicates a higher level of trust.

Respondents were given the opportunity to explain their answers to the hypothetical questions of emulating Snowden, with free text answers. At the time of writing, for most countries, including Sweden, these answers have not yet been analysed. However, the Chinese responses have been analysed. More than a third (36.5%; 42 of 115) of Chinese respondents who said they would not emulate Snowden in the PRC because of the risk not only to themselves, but their family, friends and acquaintances, due to the possibility of government reprisals [12]. A smaller but not negligible 20.9% (24 of 115) responded that they believed that state surveillance was necessary for public security in the PRC [16, 8, 2]. Chinese respondents value their privacy a great deal, perhaps partly because they are denied it, but also because they see and feel the consequences of a lack of privacy, particularly with regards to an untrusted government.

5. FURTHER WORK

These surveys represent a significant international snapshot of attitudes to privacy and surveillance across a broad range of countries. In addition to the planned deployment of these surveys in the UK and Canada, further statistical analyses of these results is expected to demonstrate other interesting factors. In particular, this paper and the others on each country have so far only looked at the results regarding Snowden's actions whereas the full survey also asked about Chelsea Manning's release of US military and diplomatic information via Wikileaks. Once the core research team has conducted their analyses, the raw survey data will be made available online for other researchers to investigate.

6. ACKNOWLEDGEMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006 and the JSPS Grant-in-Aids for Scientific Research (B) 24330127 and (B) 25285124.

In addition to the authors of this international comparison paper, the following research partners translated the survey into local languages (where necessary), arranged for respondents to take the survey and produced papers analysing their local results: Michel Schleusenet, Sarah Stevens and Sebastian Brenner (Germany); Mario Arias-Oliva (Mexico and Spain); Juan Carlos Yáñex-Luna and Pedro I. González Ramiírez (Mexico); Gehan Gunasekara (New Zealand); Jessie Duan and Dang Ronghua (the PRC and Taiwan); Ryoko Asai and Iordanis Kavatzthopoulos (Sweden).

Jessica Chai, Chen Xuan, Iheng Lirong, Jiang Xinghao, Wang Yang, Wei Yi and Zhang Ao, all students at the School of Commerce at Meiji University provided invaluable help in translating and deploying the Taiwanese and Chinese versions of the survney, and in interpreting the free text answers.

65 academics from Universities around Japan helped in encouraging their students to respond to our survey. There is no space to list them here, but the authors wish to extend their sincere thanks for their efforts.

7. REFERENCES

[1] A. A. Adams, K. Murata, and Y. Orito. The Japanese Sense of Information Privacy. *AI & Society*, 24(4):327–341, 2009.

[2] D. Bamman, B. O'Connor, and N. Smith. Censorship and deletion practices in chinese social media. *First Monday*, 17(3), 2012.

[3] D. Desilver. Most young Americans say Snowden has served the public interest. tinyurl.com/o2gmfql, 1 2014. Pew Research Fact Tank Report.

[4] D. H. Flaherty. *Protecting privacy in surveillance societies: The federal republic of Germany, Sweden, France, Canada, and the United States*. UNC Press Books, 1989.

[5] G. Greenwald. *No place to hide: Edward Snowden, the NSA, and the US surveillance state*. Metropolitan Books, 2014.

[6] M. Gurnow. *The Edward Snowden affair: Exposing the politics and media behind the NSA scandal*. Blue River Press, Indianapolis, IN, 2014.

[7] L. Harding. *The Snowden files: The inside story of the world's most wanted man*. Vintage Books, New York, NY, 2014.

[8] A. Jacobs. China Further Tightens Grip on the Internet. tinyurl.com/nrhxtty, 2015. New York Times. 29th January.

[9] P. Lundin. Computers and Welfare: The Swedish Debate on the Politics of Computerization in the 1970s and the 1980s. In C. Gram, P. Rasmussen, and S. D. Østergaard, editors, *History of Nordic Computing 4*, volume 447 of *IFIP Advances in Information and Communication Technology*, pages 3–11. 2015.

[10] E. Moglen. Privacy under attack: The NSA files revealed new threats to democracy. tinyurl.com/kcocf7o, 2014. The Guardian. 27th May.

[11] K. Murata, Y. Fukuta, Y. Orito, A. A. Adams, and A. M. Lara Palma. So What If The State Is Monitoring Us? Snowden's Revelations Have Little Social Impact in Japan. *Computers and Society*, 2015. Forthcoming.

[12] M. Nowak. Civil and political rights, including the question of torture and detention: Report of the special rapporteur on torture and other cruel, inhuman or degrading treatment or punishment, manfred nowak. tinyurl.com/ph3klfp, 3 2006.

[13] Pew Research Center. Obama's NSA Speech Has Little Impact on Skeptical Public. tinyurl.com/nw2fpfs, 2014.

[14] Pew Research Center. Public Perceptions of Privacy and Security in the Post-Snowden Era. tinyurl.com/p2536wh, 2014.

[15] L. Rainie and M. Madden. Americans' Privacy Strategies Post-Snowden. tinyurl.com/lttqx58, 2015.

[16] Reporters Without Borders. The Enemies of the Internet Special Edition: Surveillance. tinyurl.com/npctnfv, 2013.

Snowden seems to have more social impact in the People's Republic of China than in the Republic of China, but

Kiyoshi Murata
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
kmurata@meiji.ac.jp

Yasunori Fukuta
School of Commerce
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
yasufkt@meiji.ac.jp

Andrew A. Adams
Centre for Business Information Ethics
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
aaa@meiji.ac.jp

Duan Xiongfang
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
duanxiongfang@126.com

Dang Ronghua
Meiji University
1-1 Kanda Surugadai, Chiyoda,
Tokyo 101-8301, Japan
dangronghua@163.com

Ana María Lara Palma
Civil Engineering Department
Management Area
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
amlara@ubu.es

ABSTRACT

This study investigates how Snowden's revelations are viewed by young people in the People's Republic of China (PRC) and the Republic of China (Taiwan) through questionnaire surveys of and follow-up interviews with university students in those countries. Considering the history of state surveillance in both countries and the current complicated and delicate cross-strait relationships, it is interesting to examine PRC and Taiwanese youngsters' attitude and reactions to Snowden's revelations separately and in comparison.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues: abuse and crime involving computers, privacy, use/abuse of power

General Terms

Security, Human Factors, Legal Aspects

Keywords

Edward Snowden, privacy, state surveillance, social impact, the PRC, Taiwan

1. INTRODUCTION

On 13th June 2013, eight days after Edward Snowden's first revelations from Hong Kong about controversial signals intelligence carried out by the US' National Security Agency (NSA) in cooperation with intelligence agencies of other Five Eyes countries, South China Morning Post published an article including Snowden's claims that the Prism Programme included people and institutions in Hong Kong and the People's Republic of China (PRC) and that the NSA had been hacking into computers in the Special Administrative Region and on the mainland since 2009 [7]. The US government had previously strongly criticised China for hacking and surveillance outside its borders. Subsequent revelations made by Snowden included the NSA spying on PRC companies including Huawei, the world's second largest supplier of networking equipment [12] and the use of undercover operatives of the agency in global communications companies based in the PRC, Germany, Korea and even America to gain access to their sensitive data and systems [9]. These aroused the suspicion of the NSA's involvement not only in political but also in industrial espionage.

Although Snowden said "People who think I made a mistake in picking Hong Kong as a location misunderstand my intentions. I am not here to hide from justice, I am here to reveal criminality." and "The reality is that I have acted at great personal risk to help the public of the world, regardless of whether that public is American, European, or Asian." [7], the fact that he made his revelations from the territory of the PRC triggered criticism in the US that his disclosure benefited that communist country, where the circumstances surrounding state surveillance and secrecy were far worse than the US [11]. "According to a poll, China's population opposes Snowden's extradition by a significant margin, and the American has emerged as something of a folk hero in the country" [11].

Considering that the PRC is now the world's second largest economy, that the US is a strong military ally of the Republic of China (Taiwan) and that both communist and nascent democratic China politically claim the legitimacy of their

regimes whereas cross-strait economic exchanges are increasing, the attitudes of young people in the PRC and Taiwan to Snowden's revelations are of strong interest.

This study investigates how Snowden's revelations are viewed by young people in the PRC and Taiwan through questionnaire surveys of and follow-up interviews with university students in the respective countries, taking the histories of state surveillance in these countries and the current complicated and delicate cross-strait relationships into account. It is part of a larger international study covering multiple countries, and is inspired by the work of the Pew Research Center in studying US citizens' reactions to the Snowden revelations [10].

This paper first introduces the political background of both countries with respect to the issue of state surveillance and censorship. Following an overview of the survey and respondents, a detailed analysis of the most interesting results is presented. Finally, some initial conclusions are drawn about the attitudes of young people in the PRC and Taiwan with respect to issues of state surveillance of individuals by their own and foreign governments.

2. STATE SURVEILLANCE IN COMMUNIST AND NASCENT DEMOCRATIC CHINA

2.1 State Surveillance in the PRC

Since December 1987 when the "reform and opening-up" policies were adopted under the leadership of Deng Xiaoping, the PRC has attempted to move to a market economy from a planned economy system, while holding on firmly to their single-party regime. In the early 21st century, thanks partly to their entry into the World Trade Organisation, the PRC has come to play the role of "workshop of the world", and, owing to the resultant remarkable economic growth, the PRC market is now recognised as one of the most promising consumer markets in the world. On the other hand, however, the socialist market economy systems centred on economic development of the PRC coast have led to serious internal economic disparity between urban and rural residents and between the Han Chinese, who comprise 92% of the total population of the PRC, and ethnic minorities. In addition, the remarkable, but imbalanced, economic growth and the rule of man, not law, have aggravated the PRC's "traditional" corruption among bureaucrats [3]. Consequently, feelings of inequality and discontent among the public have grown in the PRC and the recent slowdown in economic growth in the country has accelerated growth of these feelings.

To repress domestic resentment and suppress pro-democracy and dissident movements and national liberation or separatist movements in Taiwan, Tibet and the Xinjiang Uighur Autonomous Region, mass state surveillance systems have been created and are operated mainly by the Ministry of State Security and the Ministry of Public Security [6]. Reflecting the historic fact that religious bodies have played a key role in previous dynastic collapses in China, participants in (learners of) the Falun Gong have also been subject to state surveillance [8] and suppression.

The widespread use of the Internet in the PRC since the early years of this century added a new dimension to internal state surveillance. The PRC government set up their Internet monitoring and censorship systems known as the Great Firewall of China, which began broad operation in 2003 [13]. It is alleged that two million government agents constantly monitor the Internet in the PRC. On the other hand, online services like Weibo are used as a "human flesh search engine" demonstrating the power of the Internet to also function as a weapon of the weak [6, 1]. In July 2006, Amnesty International [2] reported that Yahoo!, Microsoft and Google took part in Internet censorship in the PRC. The suppression of freedom of expression and information in the PRC had been regularly criticised by the US government. Ironically, however, various companies based in this democratic nation are alleged to have collaborated with the authorities of that authoritarian government. Moreover, Edward Snowden started his revelations of the true picture of state surveillance or, more precisely, surveillance conducted by an industrial-government complex of the democratic nations the US and the UK, on 5th June 2013 while in Hong Kong (part of the PRC), where he had concealed himself from US authorities while making his initial revelations.

2.2 State Surveillance in Taiwan

As the consequence of the loss of the Chinese Civil War, the Kuomintang government made a statement about their relocation from Mainland China to Taiwan on 7th December 1949. They maintained their single-party regime in Taiwan for 38 years claiming that they were the legitimate government of a unitary China. The enforcement of the Temporary Provisions Effective during the Period of Communist Rebellion order, which superseded the Constitution, in May 1948, the subsequent introduction of martial law in May 1949 and military assistance from the US in the wake of the Korean War which broke out in June 1950, allowed them to establish Kuomintang-Party as a single-party state centring on Chiang Kai-shek and his son Chiang Ching-kuo. To maintain their political grip, the Kuomintang government set up a National Security Council and an associated executive agency, the National Security Bureau (NSB), in February 1967. The NSB threw its mantle over police and secret security and intelligence agencies and kept a close watch on all Taiwanese political activities in the name of national security, as indicated by its nickname of Taiwan's KGB or TKGB.

In September 1986, the Democratic Progressive Party (DPP) was illicitly formed but eventually the Kuomintang accepted it as a legitimate opposition party, leading to the end of the single-party regime and the beginning of democratisation in Taiwan [5]. Martial law was ended in July 1987 by a presidential order issued by Chiang Ching-kuo. The Temporary Provisions Effective during the Period of Communist Rebellion order was abrogated in May 1991 under President Lee Teng-hui, who took the presidency in January 1988 and pressed on with peaceful democratisation [4]. On the other hand, the amazing economic growth centred on the export industry since the 1960s had already pushed Taiwan into a position of economic power in Asia. However, Taiwan's recent increased economic dependence on the PRC especially since the conclusion of the Economic Cooperation Framework Agreement with the PRC in June 2010 has made sitting president Ma Ying-jeou's steering of the cross-strait relationships more difficult as symbolised by the sunflower student

movement which took place in spring 2014 in opposition to the Cross-Strait Service Trade Agreement.

3. OVERVIEW OF THE SURVEYS

The questionnaire surveys of PRC and Taiwanese students were conducted using online questionnaire websites in December and October 2014, respectively. 315 of 324 responses from the PRC were valid as were all 111 responses from Taiwan. The questionnaires for these countries were developed based on the original English one created by three of the authors (Murata, Adams and Lara Palma) and were translated in collaboration with eight PRC and one Taiwanese master's course students at the Graduate School of Commerce, Meiji University.

The questionnaire used in this survey consists of three parts plus optional fact sheets. The first part was answered by all of the respondents and included questions related to right to privacy and a privacy invasion. The second part of the questionnaire was composed of questions for respondents who had already known about Snowden's revelations before the survey. In this section, respondents were requested to evaluate their recognition of and interest in Snowden's revelations. After reading a short story which gave an overview of the Snowden affair, drafted by the authors, each respondent was then asked questions relating to their evaluation of and sympathy with Snowden's activities. The questionnaire contained multiple-choice questions which allowed one answer or multiple answers plus some open-ended questions.

The male-female ratio and the age distribution of the respondents in the PRC and Taiwan are shown in Tables 1 and 2, respectively.

Table 1: Respondent attributes in the PRC (number (%))

Gender	**Male**				**Female**			
	100 (31.7%)				215 (68.3%)			
Age	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**
	34 (10.8%)	29 (9.2%)	35 (11.1%)	33 (10.5%)	37 (11.7%)	23 (7.3%)	28 (8.9%)	96 (30.5%)

Table 2: Respondent attributes in Taiwan (number (%))

Gender	**Male**				**Female**			
	45 (40.5%)				66 (59.5%)			
Age	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**
	1 (0.9%)	0 (0.0%)	9 (8.1%)	6 (5.4%)	9 (8.1%)	3 (2.7%)	35 (31.5%)	48 (43.2%)

The respondents based in the PRC and Taiwan were recruited for their participation in the questionnaire survey through personal connections with students from those countries studying at Meiji University. As part of the follow-up research to the analysis of the survey results, eight master's course students from the PRC studying at Meiji University who were not part of the respondent cohort were also interviewed in person in July 2015. Seven of them had heard about Snowden's revelations. Another eight master's course students (seven from the PRC and one from Taiwan) who were part of the respondent cohort answered follow-up questions in writing by email, also in July 2015.

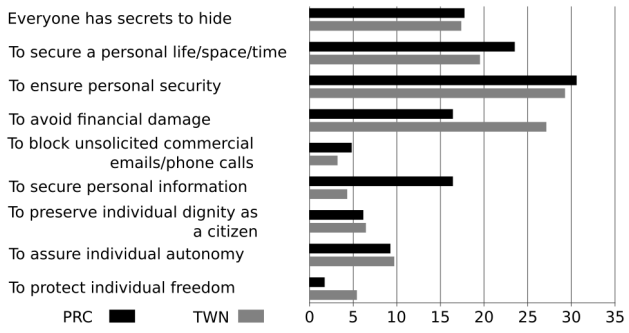


Figure 1: Why is the Right to Privacy Important?

4. SURVEY RESULTS AND DISCUSSIONS

4.1 Circumstances Related to Snowden's Revelations in the PRC and Taiwan

4.1.1 Attitude towards the Right to and an Invasion of Privacy in the PRC and Taiwan

To investigate the attitudes of respondents in the PRC and Taiwan towards Snowden's revelations in detail their perceptions of the importance of the right to privacy, their knowledge level about that right, and their attitudes towards surveillance by government and private sector organisations were examined based on the results of the survey and follow-up interviews.

The results of the survey demonstrated that both PRC and Taiwanese respondents were aware of the importance of their right to privacy. As shown in Table 3, 94.1% of PRC respondents (255 of 271) answered Q10 (Is your right to privacy important?) with "very important" (55.0%) or "important" (39.1%) and 97.2% (103 of 106) of Taiwanese respondents answered the question with "very important" (50.0%) or "important" (47.2%). The responses to open-ended question "Please describe why your right to privacy is important" (Q11) are summarised in Figure 1. Both in the PRC and Taiwan, around 30% of those respondents who considered their right to privacy was very important or important (30.7% (69 of 225) in the PRC and 29.3% (27 of 92) in Taiwan) mentioned that the right was important to ensure personal security. Whereas more than one out of four Taiwanese respondents (27.2%; 25 of 92) pointed out the connection between privacy protection and the avoidance of financial damages, only 16.4% of PRC respondents (37 of 225) did.

The degree of understanding of the right to privacy was measured using a self-esteem scale, the majority of respondents claiming to have good understanding of the right (Table 3). In Taiwan, more than eight out of ten respondents (81.2%; 82 of 101) answered that they understood the right well ("understand very well": 8.9% (9 of 101); "understand": 72.3% (73 of 101)), whereas 64.4% of respondents (163 of 253) claimed good understanding in the PRC ("understand very well": 5.9% (15 of 253); "understand": 58.5% (148 of 253)). However, in follow-up interviews, many said that they

had not learned about the right to privacy at schools, while others pointed out that they had little awareness of privacy because they were kept under Internet surveillance by the state in the PRC (except in Hong Kong).

Table 3: Awareness and Understanding of the Right to Privacy

Q10. Is your right to privacy important?		
Answer	Frequency (%)	
	PRC	Taiwan
Very important	149 (55.0%)	53 (50.0%)
Important	106 (39.1%)	50 (47.2%)
Not so important	15 (5.5%)	3 (2.8%)
Not important at all	1 (0.4%)	0 (0.0%)
Total	271	106
Q13. How well do you understand what the right to privacy is?		
Answer	Frequency (%)	
	PRC	Taiwan
Understand very well	15 (5.9%)	9 (8.9%)
Understand	148 (58.5%)	73 (72.3%)
Hardly understand	89 (35.2%)	18 (17.8%)
Don't understand at all	1 (0.4%)	1 (1.0%)
Total	253	101

According to a textual analysis of responses to the open-ended question "Please describe what the right to privacy is" (Q14), a large majority of PRC respondents who claimed they were understood the right (74.2%; 98 of 132) and a half of Taiwanese respondents who did so (50.0%; 38 of 76) considered that personal information protection was the core of the right (see Figure 2). After transforming these four-pointed scaled responses to Q10 and Q13 into two categories (Table 4), a Chi-square test was conducted to examine the relationship between the perceived importance and understanding level of the right to privacy in the PRC. The result of the test indicated there was a statistically significant positive relationship between these two variables in the country (Chi-square (1) = 15.549, $p < .01$; Phi coefficient = .248, $p < .01$). This means that those PRC respondents who felt that the right to privacy was important tended to claim a good understanding of the right and vice versa. Unfortunately, in terms of Taiwanese responses to Q10, the sample size was too small and unbalanced to perform a useful Chi-square, as shown in Table 3.

The authors examined feeling of and attitude towards a privacy invasion of PRC and Taiwanese youngsters as well. In terms of the perceived risk level of a privacy invasion, more than 80% of PRC respondents (83.2%; 257 of 309 who responded to Q6 "Do you feel that your use of the Internet involves taking risks with your privacy?") felt that their online activities involved taking risks with their privacy "strongly

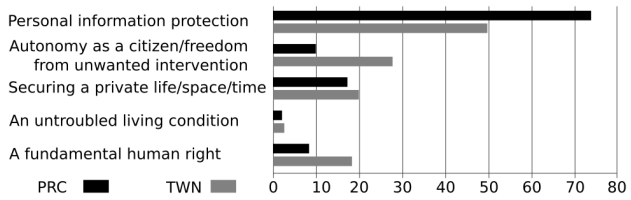


Figure 2: What Is the Right to Privacy? (%)

Table 4: Cross-tab of Responses to Q10 and Q13 in the PRC

		Q13		
		Understand	Not Understand	Total
Q10	Important	160	77	237
	Not important	3	13	16
	Total	163	90	253

(The four-point scale answers to each question were transformed into two categories. In Q10, for example, "very important" and "important" were conflated to "important".)

(20.4%)" or "to an extent (62.8%)". On the other hand, nearly 70% of them (69.9%; 216 of 309 who responded to Q7 "Do you feel that your non-Internet activity involves taking risks with your privacy?") also perceived a risk of privacy invasion associated with non-Internet activities. This result indicates that the use of the Internet had been seen as only one of the major privacy threats in the PRC. However, the result of paired samples t-test examining the statistical significance of the difference between mean scores of responses to Q6 (M = 2.02, SE = .037) and Q7 (M = 1.82, SE = .037) showed that the perceived risk of privacy invasion associated with Internet use was statistically significantly higher than the risk in the non-Internet context, at 1% significance level (D= .20, 95% CI [.117, .285], t (308) = 4.811, $p < .01$).

Furthermore, PRC respondents perceived a higher risk of a privacy invasion associated with non-Internet activities compared to respondents in other Asian countries studied: specifically, the percentage of respondents who reported feeling at risk (69.9%) was significantly higher than ones in Taiwan (53.2%; 59 of 111) and Japan (52.0%; 931 of 1792). Many of the interviewees mentioned that frequent forgery of personal identification cards, which PRC citizens were required to always carry, once they reach 16 years of age, and the resultant banking- and credit-card frauds, were seen as a major threat to privacy in the PRC.

In Taiwan, more than seven out of ten respondents (71.2%; 79 of 111) answered that their use of the Internet involved taking risks with their privacy "strongly" (2.7%; 3 of 111) or "to an extent" (68.5%; 76 of 111) while more than 50% (53.2%; 59 of 111) felt the risk in the non-Internet context "strongly"(2.7%; 3 of 111) or "to an extent" (50.5%; 56 of 111). A t-test for paired samples was carried out in order to examine whether there was a significant difference in mean scores of perceived privacy risks associated with the Internet

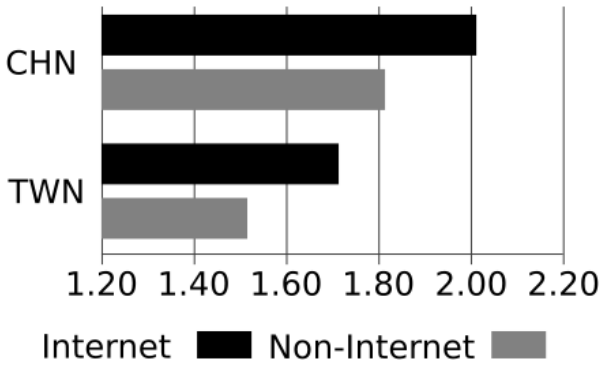


Figure 3: Ranked Means of Perceived Risks Associated with Internet and Non-Internet Activities

and non-Internet activities. The average scores of responses to Q6 and Q7 were 1.72 (SE = .05) and 1.52 (SE = .06), respectively. The difference between these averages was 0.198 (95% CI [.072, .324]) and the result of t-test indicated this difference was statistically significant at 1% significance level (t (110) = 3.242, p < .01). According to this result, it can be seen that Taiwanese youngsters regard the Internet activities as at significant risk of privacy invasion.

As shown in Table 5, the percentages of Taiwanese youngsters' perceived risks of a privacy invasion associated with Internet and non-Internet activities were less than ones of the PRC counterpart. This tendency is confirmed when ranked means of responses to Q6 and Q7 (3: strongly; 0: not at all) are calculated, as shown in Figure 3.

Table 5: Feeling of Privacy Risk with (non-)Internet Activity

	Q6. Do you feel that your use of the Internet involves taking risks with your privacy?		Q7. Do you feel that your non-Internet activity involves taking risks with your privacy?	
Answer	Freq. (%)		Freq. (%)	
	PRC	Taiwan	PRC	Taiwan
Strongly	63 (20.4%)	3 (2.7%)	40 (12.9%)	3 (2.7%)
To an extent	194 (62.8%)	76 (68.5%)	176 (57.0%)	56 (50.5%)
Not much	48 (15.5%)	30 (27.0%)	91 (29.4%)	48 (43.2%)
Not at all	4 (1.3%)	2 (1.8%)	2 (0.6%)	4 (3.6%)
Total	309	111	309	111

Results of the survey also provide the information about what kinds of organisations were/weren't viewed as threats to respondents' privacy. Tables 6 and 7 show the ranked means (3: high; 0: low) and standard deviations of responses to Q8 (How much do you feel that the following groups threaten your privacy?) in the PRC and Taiwan, respectively. Internet companies and telecom companies/Internet providers tended to be viewed as a threat to privacy by both PRC and Taiwanese respondents. Computer software companies, system integrators and other for-profit companies were also among the top-ranked in the two countries. On the other hand, the PRC respondents seemingly tended not to regard government agencies (including law enforcement agencies and secret service agencies) as a threat to privacy, whereas in Taiwan respondents considered those government agencies more risky in terms of an invasion of their privacy. In follow-up interviews, almost everyone suggested that in the PRC everyone supposed his/her personal information was held by police agencies, but not misused by them, while for-profit companies would not hesitate to misuse personal information for reaping profits. Ordinary Chinese, the interviewees said, did not need to worry about police. It was also pointed out during the interviews that educational institutions could be considered as a threat to privacy because it was not unusual in the PRC for high schools to sell the contact information of their students to three-year occupational colleges so that they could directly send college enrolment information to students (there is intense competition between the colleges in the PRC).

Table 6: Ranked means (0:low; 3: high) of 15 groups as perceived privacy threat (PRC)

Q8. How much do you feel that the following groups threaten your privacy?		
Group	Mean	S.D.
Internet companies	2.48	.682
Telecom companies/Internet providers	2.40	.738
Other for-profit companies	2.03	.803
Computer software companies	1.90	.823
System Integrators	1.89	.849
Educational institutions	1.87	.817
Computer hardware companies	1.78	.850
Individuals who you don't know	1.68	.815
Individuals who you know but not well	1.68	.637
Health-care organisations	1.59	.823
Other not-for-profit organisations	1.49	.783
Other government agencies	1.37	.868
Secret service government agencies	1.37	.931
Law enforcement government agencies	1.32	.902
Individuals who you know well	1.32	.775

4.1.2 The Degree of Recognition of and Interest in Snowden's Revelation in the PRC and Taiwan

The percentages of respondents who had heard about Snowden's revelations were different between the PRC and Taiwan. Whereas more than three out of four PRC respondents had heard the revelations before the questionnaire survey (76.4%; 188 of 246 who responded to Q19 "Have you heard about Snowden's revelations?"), Taiwanese respondents who had heard about the revelations were a bare minority (46.5%; 47 of 101), perhaps reflecting Snowden's presence in Hong Kong when he started his revelations and that the affair was highly publicised in the PRC. The interviewees admitted that TV, newspapers and Internet news sites repeatedly reported the Snowden affair as America's failure for at least for three months after his first revelations.

Table 7: Ranked means (0:low; 3: high) of 15 groups as perceived privacy threat (Taiwan)

Q8. How much do you feel that the following groups threaten your privacy?		
Group	Mean	S.D.
Internet companies	2.35	.604
Telecom companies/Internet providers	2.11	.652
System Integrators	2.01	.692
Other for-profit companies	1.98	.718
Secret service government agencies	1.82	.780
Computer software companies	1.76	.701
Individuals who you don't know	1.69	.817
Other government agencies	1.67	.779
Law enforcement government agencies	1.67	.836
Computer hardware companies	1.64	.686
Educational institutions	1.59	.743
Health-care organisations	1.58	.706
Other not-for-profit organisations	1.51	.680
Individuals who you know but not well	1.48	.639
Individuals who you know well	1.19	.741

Survey results also reveal respondents' self-report of their degree of understanding of Snowden's revelations in the two countries. Those respondents who had heard about the revelations before the survey were requested to evaluate the level of their understanding of the Snowden affair in the following three dimensions: the contents of Snowden's revelations, the US government's reactions to them and the current status of Snowden. 53.3% of PRC youngsters (99 of 186 who responded to Q23 "How much do you know about the contents of Snowden's revelations?") knew the contents of the revelations "a lot" (1.1%) or "a fair amount" (52.2%). In Taiwan, 48.9% of respondents (23 of 47) answered the same question with "a lot" (2.1%) or "a fair amount" (46.8%). Likewise, to the question about the US government's reactions and Snowden's current status, 47.3% (90 of 187) and 28.3% (53 of 187) of PRC respondents answered "a lot" and "a fair amount" respectively, but only 42.6% (20 of 47) and 12.7% (6 of 47) of Taiwanese respondents..

Respondents were asked about their level of interest in the revelations via two questions: Q21 (Have you ever talked about Snowden's revelations with others?) and Q22 (Have you ever searched for information about Snowden's revelations?). 42.2% of PRC respondents (79 of 187) had talked about the revelations with others and 45.7% of them (86 of 188) had searched for information. Meanwhile, in Taiwan, 27.7% (13 of 47) of respondents had discussed the revelations with others and 30.4% (14 of 46) had searched for information.

These survey results seem to indicate that in general youngsters living in the PRC are interested in Snowden's revelations and know them well. On the other hand, Taiwanese youngsters' degree of interest in and knowledge of the revelations are below that of those in the PRC.

4.1.3 Evaluation of Snowden's Activities in the PRC and Taiwan

Respondents' evaluation of Snowden's revelations were also sought. To a question about the social contribution of the revelations (Q28: Have Snowden's revelations served the public interest or harmed it?), around one out of four respondents, more specifically 59 of 230 (25.7%) in the PRC and 26 of 98 (26.5%) of Taiwanese respondents avoided making a clear judgement answering it with "no option" or "prefer not to answer". Amongst those respondents who offered a judgement on whether Snowden served the public interest or not, 84.2% of PRC respondents (144 of 171 who indicated Snowden's revelations served the public interest "a lot" (15.8%) or "to an extent" (68.4%)) and 79.2% of Taiwanese youngsters (57 of 72 responded "a lot" (7.0%) or "to an extent" (72.2%)), positively evaluating the revelations. Even when the respondents who answered Q28 with "no option" or "prefer not to answer" being taken into account, 62.6% of respondents in the PRC (144 of 230) and 58.2% in Taiwan (57 of 98) clearly gave a positive evaluation to the Snowden revelations in terms of public interest. Many of the follow-up interviewees mentioned that the press coverage of the Snowden affair in the PRC was favourable to him, condemning the hypocrisy of the US government prior criticisms PRC government's control of information. Responses to the open-ended question "Why do you think Snowden determined to make those revelations?" (Q27) demonstrate that more than a half of Taiwanese respondents (51.9%; 40 of 77) and nearly 40% of PRC respondents (37.5%; 69 of 184) considered Snowden decided to made the disclosure based on his criticism against the surveillance and privacy invasion by the government agencies (Figure 4), with very few attributing baser motives (self-protection, general anti-American sentiment or being an agent of a foreign power).

4.2 Empirical Consideration about Influence of Snowden's Revelations in the PRC and Taiwan

4.2.1 The Impact of Snowden's Revelations on Society

The effects of Snowden's revelations were examined through analysing the responses to the open-ended question Q36 (What social changes do you think have happened because of Snowden's revelations?). About 40% of PRC and Taiwanese respondents (41.0% (93 of 227) and 38.9% (37 of 95), respectively) were able to cite an instance of social change led by Snowden's revelations, whereas the ratio of the respondents who judged the revelations had not created any social change was less than 5% in the PRC (4.0%; 9 of 227) and only 2.1% (2 of 95) in Taiwan. However, attention is drawn to the fact that 36.1% of PRC respondents (82 of 227) and 42.1% of Taiwanese (40 of 95) offered no opinion about changes caused by the revelations.

Amongst those who mentioned some sort of social changes caused by the revelations, 49.5% (46 of 93) of PRC respondents and 45.9% (17 of 37) of Taiwanese respondents considered people's awareness of privacy had been enhanced and 37.6% (35 of 93) and 32.4 % (12 of 37) of respondents in the PRC and Taiwan, respectively, felt the trust in the American governments had been eroded. Whereas 48.6% (18 of

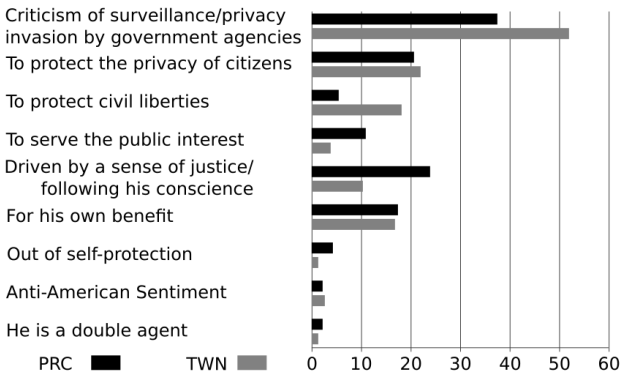


Figure 4: Why Did Snowden Determine to Make the Revelations? (%)

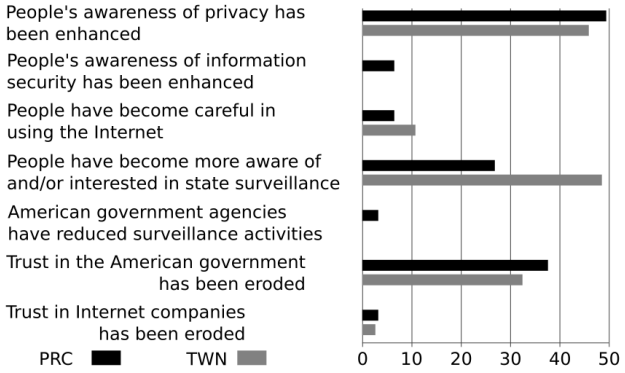


Figure 5: What Social Changes Have Happened Because of Snowden's Revelations? (%)

37) of respondents in Taiwan believed people had become more aware of and/or interested in state surveillance, 26.9% (25 of 93) in the PRC did (Figure 5).

4.2.2 The Influence of Snowden's Revelations over the Perceived Risk of a Privacy Invasion

In order to consider the influence of Snowden's revelations over perceived risk of privacy invasion, the respondents were divided into two groups based on their response to Q19 (Have you heard about Snowden's revelations?): the group of respondents who had heard about Snowden's revelations ("Heard" group) and of those who had not ("Not heard" group). The differences in the degree of perceived risk of invasion of privacy between the groups were inspected using their ranked means of responses (3: high; 0: low) to Q6 (Do you feel that your use of the Internet involves taking risks with your privacy?), Q8-m (How much do you feel that law enforcement government agencies threaten your privacy?), Q8-n (How much do you feel that secret service government agencies threaten your privacy?) and Q8-o (How much do you feel that other government agencies threaten your privacy?) and subjected to t-tests.

If Snowden's revelations have an influence over people's perceived risk of an invasion of privacy, it is expected that the ranked means of the "Heard" group would be higher than

Table 8: Perceived Threat to Privacy from Law Enforcement Government Agencies

PRC

Heard: M=1.30, SE=.070; Not Heard: M=1.32, SE=.131; D=-.026, 95% CI [-.307, .281]; t(217) = -.176; $p > .1$

Taiwan

Heard: M=1.69, SE=.134; Not Heard: M=1.67, SE=.120; D=.022, 95% CI [-.316, .375]; t(91) = .124; $p > .1$

Table 9: Perceived Threat to Privacy from Secret Service Government Agencies

PRC

Heard: M=1.41, SE=.074; Not Heard: M=1.25, SE=.129; D=.164, 95% CI [-.121, .451]; t(217) = 1.096; $p > .1$

Taiwan

Heard: M=1.91, SE=.126; Not Heard: M=1.79, SE=.107; D=.119, 95% CI [-.199, .422]; t(91) = .724; $p > .1$

those of the "Not heard" group. However, the results of the t-tests could not show the existence of any such influence. The t-test conducted in order to estimate the relationship between Q6 and Q19 in the PRC indicated that, contrary to the expectations, the mean of the "Heard" group (M = 2.01, SE = .051) was below that of the "Not heard" group (M = 2.10, SE = .068), but the difference between these averages (D = -.093, 95% CI [-.257, .074]) was not statistically significant (t (244) = -.932, p > .1). Moreover, a t-test applied to the Taiwanese dataset showed the similar results to the PRC case. That is, the mean of the "Heard" group (M = 1.68, SE = .097) was below that of the "Not heard" group (M = 1.74, SE = .060), but the difference between the means (D = -.060, 95% CI [-.269, .144]) was not statistically significant (t (78.35) = -.526, p >.1). These results indicates that respondents had perceptions of privacy risks regardless of whether they had heard about Snowden's revelations or not, demonstrating that the revelations seem to have had no influence over the perceive risk level in the PRC or Taiwan.

Another series of t-tests were carried out to examine whether respondents of the "Heard" group felt a higher level of threat from government agencies than that of the "Not heard" group. The results are shown in Tables 8, 9 and 10. In terms of secret service government agencies (Table 9) and other government agencies (Table 10), PRC youngsters of the "Heard" group had higher average scores than those in the "Not heard" group, whereas in terms of law enforcement government agencies (Table 8) the average score of the "Heard" group was lower than those in the "Not heard" group. However, the test results indicated that these differences in the pair of means were not statistically significant. The t-tests applied to the Taiwanese dataset showed similar results. While Taiwanese respondents of the "Heard" group had higher average scores as to all the three types of government agencies than ones of the "Not heard" group, the differences in the pairs of means were not statistically significant. This indicates that the degree of perceived privacy risk from government agencies was also not influenced by whether they had heard about Snowden's revelations or not.

The results of the series of t-tests consistently showed that knowledge of Snowden's revelations had no significant in-

Table 10: Perceived Threat to Privacy from Other Government Agencies

PRC

Heard: M=1.37, SE=.068; Not Heard: M=1.32, SE=.126; D=.053, 95% CI [-.217, .328]; t(217) = .377; $p > .1$

Taiwan

Heard: M=1.73, SE=.121; Not Heard: M=1.63, SE=.114; D=.108, 95% CI [-.189, .450]; t(91) = .654; $p > .1$

fluence over respondents' perceived risk of privacy invasion. A majority of follow-up interviewees had not changed their way of using the Internet, although all the respondents said that their awareness of privacy had been enhanced because of hearing about Snowden's revelations. Only three interviewees had deleted some of their posts and refrained from making new posts on Chinese instant message service Tencent QQ.

4.2.3 *Changes in Online Communication Due to Snowden's Revelations*

Whether knowledge of Snowden's revelations produced some sort of change at an action level was also investigated. Amongst the respondents who had heard about Snowden's revelations, 45.1% of PRC respondents (73 of 162) and 52.2% of Taiwanese respondents (24 of 46) answered Q24 (Have you changed your way of communicating online using systems such as social media (e.g., Twitter, Facebook), Messenger, YouTube, blogging, Skype, email and instant messaging since you heard about Snowden's revelations?) with "have not changed at all". In other words, 54.9% and 47.8% of PRC and Taiwanese respondents of the "Heard" group, respectively, had made some change to their ways of communicating online. Even though it is substantially difficult to correctly judge the meanings of these percentages, nevertheless a significant number of PRC and Taiwanese youngsters who had got word of Snowden's revelations reported making a change in their ways of communicating online. These percentages were close to those from the parallel studies in the European countries Spain, Germany and Sweden.

4.2.4 *The Potential Influence of Snowden's Revelations over Societies*

Whether respondents would follow Snowden's lead or not when hypothetically placed in a similar situation is considered to be another indicator of the potential influence of Snowden's revelations, because such intention can be seen as predictors of acceptance of and sympathy with Snowden's behaviour. These intentions were measured by Q30 (If you were an American citizen and were faced with a similar situation to Snowden, do you think you would do what he did?) and Q33 (If you were faced with a similar situation to Snowden in your country, i.e. you found out that an intelligence agency of your country was conducting similar operations to those of the NSA and GCHQ, would you, as a Japanese citizen, do what he did?).

Interestingly, whereas a large majority of respondents recognised that Snowden's revelations had served the public interest at least to an extent in the two countries (84.2% (144 of 171) in the PRC and 79.2% (57 of 72) in Taiwan), the PRC respondents seemed very hesitant to follow Snowden's lead.

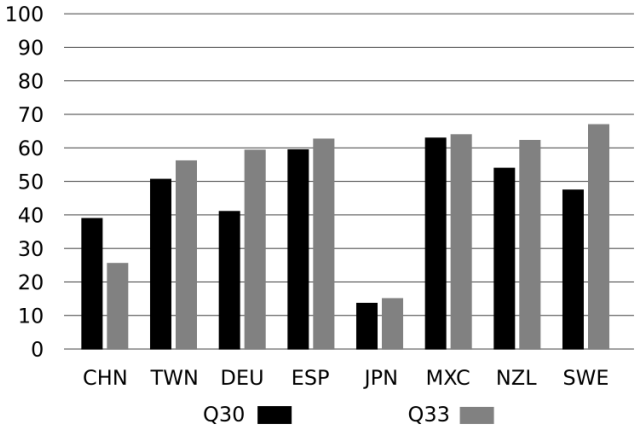


Figure 6: Percentages of "yes" to Q30/Q33 in Eight Countries

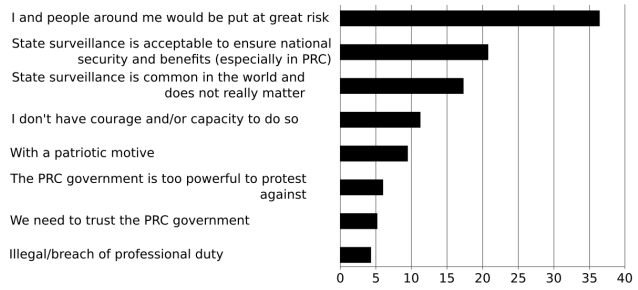


Figure 7: Why PRC Youngsters Would Not Follow Snowden's Lead in Their Country?

39.2% (69 of 176) and 50.9% (28 of 55) of respondents answered Q30 with "yes" in the PRC and Taiwan, respectively. In addition, very few PRC respondents answered Q33 with "yes" (25.8%; 47 of 182) while in contrast a similar bare majority of Taiwanese respondents did (56.4%; 31 of 55). In terms both of Q30 and Q33, the degree of PRC youngsters' intention to emulate Snowden was below that of their Taiwanese counterparts (a significant number of Taiwanese respondents preferred not to answer Q30 (42 of 97) and Q33 (41 of 96) with these non-responses treated as missing values in the above analysis). Moreover, amongst the eight countries where the surveys of this study were conducted, only in the PRC did the number of respondents who would follow Snowden's lead in the US exceed the case of their own country (Figure 6). The responses to the open-ended question about why they would not follow Snowden's lead in the PRC (Q35) more than 35% of them (36.5%; 42 of 115) considered following Snowden would put them and their family, friends and acquaintances at a great risk, including the threat to lives on the one hand, while more than 20% (20.9%; 24 of 115) believed state surveillance should be accepted to ensure societal security and benefits in the PRC considering the current situations in the country (Figure 7).

These survey results seem to reveal that Taiwanese society is more receptive to Snowden's activities than that of the PRC, and that the potential influence of Snowden's revelations in Taiwan may be greater than in the PRC.

5. CONCLUSIONS

Though both the two states investigated in this study have a country name including "China", significant differences in the social impact of Snowden's revelations were found. Simple tabulations of responses to the questionnaire used in this survey seemingly show that Snowden had more social impact in the PRC than in Taiwan. However, detailed statistical analysis demonstrates that Taiwanese respondents were more influenced by Snowden's Revelations than PRC respondents especially when actions as opposed to mere evaluation are considered.

6. ACKNOWLEDGMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006 and the JSPS Grant-in-Aid for Scientific Research (B) 25285124 and (B) 24330127. The authors appreciate the cooperation for developing the questionnaire and conducting the questionnaire surveys provided by Jessica Chai, Chen Xuan, Iheng Lirong, Jiang Xinghao, Wang Yang, Wei Yi and Zhang Ao.

7. REFERENCES

[1] T. Ako. *The Nation Battening the Poor: The Chinese Unequal Society*. Shichosha, Tokyo, 2014. In Japanese.

[2] Amnesty International. Undermining freedom of expression in China. `www.reports-and-materials.org/Amnesty-UK-report-Internet-cos-China-Jul-2006.pdf`, 2006.

[3] Q. He. *China's Pitfalls*. Shoshisa, Tokyo, 2013. In Japanese. Translated by Sakai, S. and Nakagawa, T.

[4] H. Ijiri. *The History of Politics and Foreign Diplomacy in Taiwan: Lee Teng-hui, Chen Shui-bian and Ma Ying-jeou*. Minerva Shobo, Kyoto, 2013. In Japanese.

[5] K. Ito. *Taiwan: 400-Year History and the Future*. Chuokoron-Shinsha, Tokyo, 1993. In Japanese. English translation available at `members.shaw.ca/leksu/index.htm`.

[6] R. Kashihara. *Chinese Intelligence Agencies: Spying in the World*. Shodensha, Tokyo, 2013. In Japanese.

[7] L. Lam. Edward Snowden: US government has been hacking Hong Kong and China for years. *South China Morning Post*, 2013. 13th June.

[8] Z. Li. *The Cruel Nation*. Fusosha, Tokyo, 2015. In Japanese.

[9] P. Maass and L. Poitras. Core secrets: NSA saboteurs in China and Germany. *The Intercept*, 2014. 11th October.

[10] Pew Research Center. Obama's NSA Speech Has Little Impact on Skeptical Public. `tinyurl.com/nw2fpfs`, 2014.

[11] M. Schiavenza. Edward Snowden: China's useful idiot? *The Atlantic*, 2013. 17th June.

[12] J. Shieber. Latest Snowden revelations: NSA hacks Huawei. *TechCrunch*, 2014. 23rd March.

[13] G. Walton. China's Golden Shield: Corporations and the Development of Surveillance Technology in the People's Republic of China. `publications.gc.ca/collections/Collection/E84-7-2001E.pdf`, 2001.

Snowden's revelations led to more informed and shocked German citizens

Michael Schleusener
Hochschule Niederrhein
Reinarzstraße 49, 47805
Krefeld, Germany
+49 2151 822 6610
michael.schleusener@hs-niederrhein.de

Sarah Stevens
eWeb Research Center
Webschulstraße 31, 41065
Mönchengladbach, Germany
+49 2161 186 6124
sarah.stevens@hs-niederrhein.de

Sebastian Brenner
eWeb Research Center
Webschulstraße 31, 41065
Mönchengladbach, Germany
+49 2161 186 6124
sebastian.brenner@hs-niederrhein.de

Kiyoshi Murata
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2165
kmurata@meiji.ac.jp

Andrew A. Adams
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2329
aaa@meiji.ac.jp

Ana María Lara Palma
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
+34 947 259 360
amlara@ubu.es

ABSTRACT
This study investigates the attitudes towards and social impacts of Edward Snowden's revelations in Germany through a questionnaire survey with German youngsters as part of the worldwide cross-cultural analyses. However due to Snowden‘s revelations a continuing discussion about privacy, safety, security and data protection was unleashed in Germany. The results show interesting values and settings of young people in Germany. For example: The majority 69.41% of the surveyed persons feel that the usage of the Internet threatens their privacy.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues – *abuse and crime involving computers, privacy, use/abuse of power*

General Terms
Security, Human Factors, Legal Aspects

Keywords
Snowden, Privacy, Freedom, Germany

1. INTRODUCTION AND CURRENT SITUATION IN GERMANY
The revelations of Edward Snowden have been in the German daily press since the beginning of June 2013. On the 24th of October 2013 it was revealed, that also the mobile phone of Chancellor Merkel was tapped and the public discussion started again [1]. The Federal Foreign Minister Westerwelle summoned the US Ambassador in an unprecedented action to show displeasure [2]. A discussion about a No Spy agreement was underway, even though it was not clear if the real reasons of discussion were the election and the following coalition talks. The German Federal Government does not regard Snowden's actions as politically motivated, but nevertheless would have to turn him in to the US due to bilateral agreements [3]. Therefore Germany denied Snowden's request for asylum (Rosenbach; Stark, 2014). Even until today different NGO's are seeking asylum for Snowden in Germany [4]. In March 2014, all parties of the Bundestag agreed to form a parliamentary board of enquiry in regard to the NSA in a rare occasion of joint unity [5].

Current situation in Germany (May 2015): As early as 2011 Snowden said that the BND and the NSA worked more closely together than known at that time. In March 2015 it was revealed that in Bad Aibling, Germany, a big large scale bugging base of the US Secret Services has been in operation. Two or three times a day the BND picked up search terms from an American server. According to information by "DER SPIEGEL" [1] there could be a total of as many as 40 000 search terms. These terms have been used for searching in private data like e-mail- and phone-communication of German citizens. In case of technical problems the NSA has been consulted repeatedly [6]. In April 2015 the

[1] "DER SPIEGEL" is a major German news magazine.

Federal Chancellery gave instructions to eliminate the organizational shortcomings inside the BND [7].

However due to Snowden's revelations a continuing discussion about privacy, safety, security and data protection was unleashed in Germany. The discussion expands to data collection and data analysis of social networks, big search engines and to generating meta-information while using smartphones as well as applications running on them.

2. OVERVIEW OF THE SURVEY

This study deals with the attitudes and social impacts in Germany. 81 valid responses have been collected from November 2014 to January 2015. Most of the persons surveyed were German citizens (92 %) and they show a typical gender distribution of 53 % female and 47 % male persons. Up to 84 % were students in a bachelor-degree and not more than 30 years old.

The age spread of responders is weight towards young-age: 28.4 % (23/81) of the responders are 18-20 years old, 40.7 % (33/81) between 21 and 24 years old and 30.9 % (25/81) are at least 25 years old or older.

A majority of 86.3 % (69/80) of the respondents are currently studying. 6.3 % (5/80) are working and 7.5 % (6/80) are working and studying in a dual education system.

The two universities with the highest amount of respondents who are studying currently at their facilities are Hochschule Niederrhein with 45.6 % (31/68) and Hochschule Fresenius with 39.7 % (27/68). The rest is made up of 5.9 % (4/68) HMKW and 8.9 % (6/68) from other universities of the regional area.

The sample has a good spread which makes it possible to differentiate the following outcomes according to the primary area of study. In detail 12.8 % (10/78) study Humanities, 20.5 % (16/78) Engineering, 19.2 % (15/78) Economies, 19.2 % (15/78) Psychology, 19.2 % (15/78) Industrial Engineering, 6.4 % (5/78) Social Sciences and 9.4 % (12/78) are packed together as others, consisting of four more disciplines.

The majority of 91.4 % (74/81) in this survey are German respondents. The other 8.6 % (7/81) consist of Russian, Turkish and other nationalities.

Table 1. Respondent attributes (number of respondents (%))

Gender	**Male**		**Female**
	38 (47%)		43 (53%)
Age	**18-20**	**21-24**	**25+**
	28,4% (23)	40,7% (33)	30,9% (25)

3. SURVEY RESULTS AND DISCUSSIONS

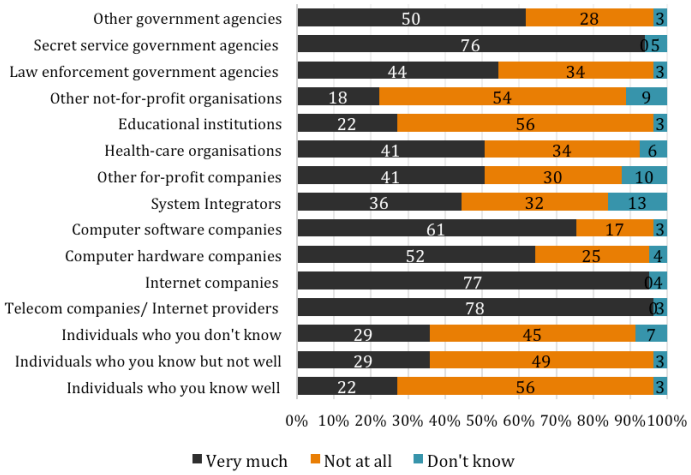
3.1 Germans are well aware threats of their privacy

The surveyed persons answer the question: "Do you feel that your use of the Internet involves taking risks with your privacy?" 93.8% (76/81) with yes and 6.2% (5/81) with no. This shows that the respondents are well aware of the threats of the Internet towards their privacy. It could be that the revelations are connected to this or that German respondents tend to be more concerned about their privacy.

3.2 Germans feel threatened by companies

As figure 2 shows internet companies with 95.1 % (77/81), secret service government agencies with 93.8 % (76/81) and telecom companies with 96.3 % (78/81) are the leading groups in terms of privacy threats according to the respondents of this survey. On the other hand individuals, non-profit organisations and educational institutions are the groups that gain the highest trust of the respondents in terms of privacy. Overall you can see that there is a negative trend towards technical products and services, as they are in the focus of many privacy discussions.

Figure 2. Germans threat of privacy (number of respondents (%))



3.3 The right of privacy is important

The right of privacy is extremely important to the respondents as you can see in Figure 3. In total 90.00 % (70/78) choose "very important" or "important" answering the question "Is your right to privacy important?" None answered "not important at all" in this case. Furthermore, the large number of 63 valid free-text responses shows the importance of privacy. Frequent formulations are "afraid to be a naked citizen", "safety is an important feeling", "fundamental right", "personal freedom" and "freedom of choice". On the other hand there were six free-text answers to the question, explaining why the right of privacy is not important. The most characteristic ones were: "I see my privacy is already lost." And "I do not threaten my privacy." Noteworthy is that the first sentence includes surrender.

Figure 3. Privacy is important (number of respondents (%))

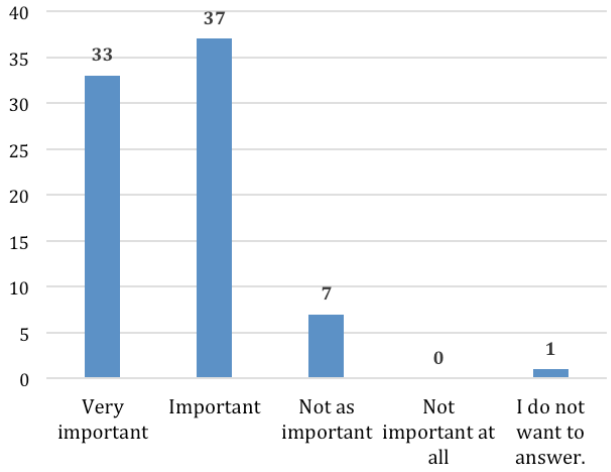
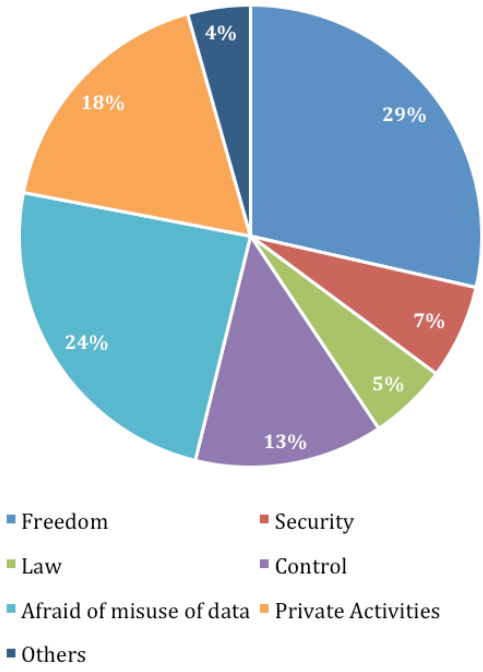


Figure 4. Reasons for important privacy (number of respondents (%))



3.4 Reasons for important privacy

The question why the right of privacy is so important is a very difficult question, as privacy and the level of attention it should get is not the same for everyone. Therefore we put the free text answers into six different groups. The first group consists of respondents who linked privacy directly to "freedom" and said that they do not want to be observed in their lives. This group shows the highest approval with 29 %. The second group said that they feel more "secure" when their right to privacy is given, 7 % see it that way. 5 % of the respondents said that the right to privacy is so important because it is in the "law" and should be available for everyone. 13 % want to maintain "control" over their personal data. The fifth group feared the "consequences" if their right to privacy is not given. A total of 24 % are afraid of companies and others who could use their data. The last group includes 18% of the respondents and which says that they want to keep their "private activities" private and that no one else should know about this. All remaining answers are combined in "others" with 4 %.

3.5 Privacy is freedom and control

Again the answers were summarized into the same six groups as in 3.4. The two biggest groups are "freedom" with 32 % and "control" with 27%. Compared to 3.4 almost every group has nearly the same size, just "control" gained 14 % and "afraid of misuse of data" lost 15 %. This could be due to a misunderstanding of the question or it could show that both groups overlap in part of given answers. On the other hand it could also mean that the misuse of data is not really an important factor in the primary definition of the right to privacy.

3.6 Germans recognize NSA and BND[2] highly

Figure 5 shows that on the one hand agencies, which were connected to the revelations, like the NSA and the BND, have been highly recognized. On the other hand agencies that play their role in movies and international news like the FBI and CIA are also well known. The GCHQ has the lowest number of respondents who know about it. But the MAD and BSI, which are both related to the German government, are also not really well known.

[2] BND = Bundesnachrichtendienst is the foreign intelligence agency of Germany

Figure 5. Knowledge of agency (%))

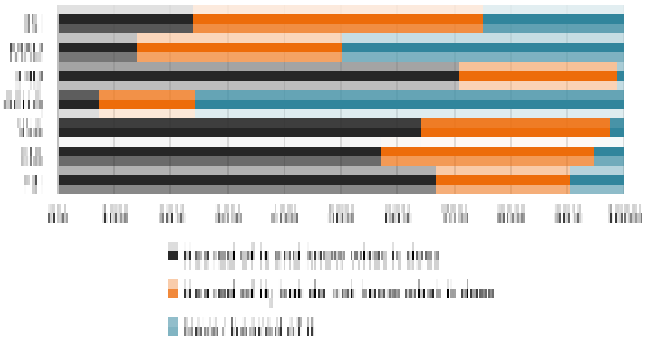


Figure 6. Reason for behaving like Snowden in Germany (%))

3.7 Snowden is better known than Manning

The Knowledge of Manning is pretty evenly split and half of the respondents have heard about it and half have not. In detail 51.4% (37/72) said yes and 48.6% (35/72) said no. In comparison to the knowledge of Snowden: 98.6 % (72/73) of the respondents have heard about Edward Snowden and only 1.4 % (1/73) have not. This statistic shows that Snowden's case had a broad audience in Germany and nearly everyone knows about him.

In this connection Germans not feels well informed about both revelations: A total of 63.1 % admitted not knowing much about the Manning's revelations. They can be split into 44.7 % (17/38) not knowing much and 18.4 % (7/38), who have only briefly heard about it. On the other hand 10.5 % (4/38) claimed to have heard a lot about it and 26.3 % (10/38) at least claimed to know a fair amount. A high figure of 91.7 % say they do not know very much about Snowden's revelations. In detail the results can be read out as followed: 0.0 % (0/72) claimed to know a lot, 8.3 % (6/72) a fair amount, 54.2 % (39/72) not much and 37.5 % (27/72) said they just know little about the revelations.

But 73.6 % (53/72) of the respondents answered the question "Have you ever talked about Snowden's revelations with others?" with yes and 26.4 % (19/72) with no. It seems to be a topic German students talk about.

As Germany is known as one of the great supporters of Snowden it is no surprise that 86.4 % (57/66) of the respondents said that his revelations served the public interest and only 13.6 % (9/66) said it harmed it. As in the questions before German respondents show that they sympathize with Snowden and the choices he made. So 11.1 % (6/54) said that the US government should pursue a criminal case against him, whereas 88.9 % (48/54) said it should not.

3.8 Germans would behave like Snowden

59.6 % (28/47) of the respondents answer the question "If you were faced with a similar situation to Snowden in Germany, i.e. you found out that a German intelligence agency was conducting similar operations to those of the NSA and GCHQ, would you, as a German citizen, do what he did?" yes and 40.4 % (19/47) said no. But only as a German citizen. In case of an American citizen Germans would not behave like Snowden: 41.3 % (19/46) of the respondents answered with yes and 58.7 % (27/46) answered with no.

The reason for Behaving like Snowden in Germany is to feel saver in Germany. The reason for doing not is fear.

3.9 No changes in Germany

Germans think mostly, that there are no social changes have happened because of Snowden's revelations: 40.4 % (23/57) say that nothing changed and 59.6 % (34/57) say that at least something changed. Of those 59.6 % almost everybody said that people are now more careful with their use of data. Another point that changed for some people is the relationship between Germany and the USA.

62.7 % (42/67) of the respondents think that safety and security of the society goes hand in hand with the loss of privacy and freedom. Of those, 7.5 % (5/67) think that you have to give up a lot of your freedom and privacy to ensure safety and 55.2 % (37/67) think that a fair amount is enough. On the other hand 36.8 % (25/67) respondents think that it is possible to maintain safety and security without giving up on privacy and freedom. They can be split into two groups: 26.9 % (18/67), who say you do not have to give up much and 10.4 % (7/67), who say you do not have to make any compromises.

4. CONCLUSIONS

Snowden's revelations led to more informed German citizens, but also to more shocked German citizens. The more information about the practices of secret services makes Germans more uncertain. For the same time Germans think that all these revelations of Snowden and also of Manning will not change something in their social life, but they think that the relationship between Germany and the USA may change in some way. Even When Germany repeatedly emphasized that there is a big friendship between Germany and the US, because of their aid after the Second World War. At least it is interesting to know, that the German Privacy Laws are one of the strictest Privacy Laws all over the world. After the revelations a lot of German citizens change their behavior in handling of personal data: 5.6 % (7/89) stopped using some services, 17.5 % (22/89) tried to cut down the usage of some services, 15.1 % (19/89) deleted some previously posted personal data, 18.3 % (23/89) paid more attention to which kind of personal data to publish and 22.2 % (28/89) have changed their privacy settings on some systems. Those statistics show that the German respondents are not willing to give up their beloved services but try to modify them in a way that they become

compatible with the students' changed concerns about privacy issues.

5. ACKNOWLEDGMENTS

The German authors would like to say thank you for having the possibility to participate in the international survey of privacy. A special thanks goes to Prof. Kiyoshi Murata of the Meiji University, Tokyo, as leader of the study. Also special thanks to Ana María Lara Palma and Andrew A. Adams and all other participant in the other countries.

6. REFERENCES

[1] Rosenbach M., Stark H. (2014) *Der NSA-Komplex: Edward Snowden und der Weg in die totale Überwachung* (*translation: The NSA complex: Edward Snowden and the way into total surveillance*) Spiegel Verlag Hamburg

[2] Bierling S. (2014) Vormacht wider Willen: *Deutsche Außenpolitik von der Wiedervereinigung bis ... (translation: Supremacy against his will: German foreign policy from reunification up to …*) Google eBook

[3] Gazeas N. (2014) *Deutschland müsste Snowden nicht an die USA ausliefern (translation: Germany would not extradite Snowden to the US)* Gastbeitrag in daily press "Die Zeit"Tavel, P. 2007. *Modeling and Simulation Design*. AK Peters Ltd., Natick, MA.

[4] Werkner I., et al. (2014) Friedensgutachten 2014: des Bonn International Center for Conversio LIT Verlag Münster

[5] German Bundestag Drucksache 18/483 12th February 2014

[6] Goetz J., Leyendecker H., Mascolo G., Berlin

[7] Dehmer D., Haselberger S. in the German newspaper "*Der Tagesspiegel*", 04.05.2015

Information surveillance by Governments: Impacts of Snowden's revelations in Spain

Mario Arias Oliva
Rovira i Virgili University
Av. Universitat, 1
43204, Reus, Spain
mario.arias@urv.cat

Ana María Lara Palma
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
+34 947 259 360
amlara@ubu.es

Kiyoshi Murata
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2165
kmurata@meiji.ac.jp

Andrew A. Adams
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2329
aaa@meiji.ac.jp

ABSTRACT
This study investigates the attitudes towards and social impacts of Edward Snowden's revelations in Spain through a questionnaire survey answered by students in two Spanish universities (Universitat Rovira i Virgili and Burgos University). It is part of the worldwide cross-cultural analyses about privacy perceptions in young people. The survey results take into socio-cultural and political environment.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Public Policy Issues – *abuse and crime involving computers, privacy, use/abuse of power*

General Terms
Security, Human Factors, Legal Aspects

Keywords
Edward Snowden, privacy, state surveillance, social impact, Spain

1. INTRODUCTION
Privacy and security have become a primarily concern after Snowden´s Revelations. Objectives pursuit by leaking and filtering secret information, sensitive or classified documents and their consequences have caused tiny but interesting attitudes at youngest in Spain. Globalization and technology have been a solid support for evolution, but, inherently to them, some problems arise, such as personal data life exposure, among others.

Conference'10, Month 1–2, 2010, City, State, Country.

Although very recent too, governments have set up surveillance programs but population continuing having a feeling of insecurity and slackness. The purpose of this paper is to analyze the impacts Snowden´s Revelations has had in the attitudes of our Spanish youngsters.

The structure of the remainder of this paper is as follows. There is a brief framework with the cultural, political and historical background of Spain circumstances surrounding the Snowden's revelations, a historical status of surveillance conducted by the state and the current public/people's acceptance and legal status of surveillance by the state. Continuing with the empirical research, there is an overview of the Spanish surveys and discussion about Spanish circumstances related to Snowden´s Revelations and an empirical consideration about influence of Snowden´s revelations. Results have provided interesting findings about Spanish youngest attitudes that will be point out as conclusions.

2. EVOLUTION OF STATE OF SURVEILLANCE IN SPAIN
In this section we will analyse the evolution of social and political Spanish context [1]. We consider that it is very important to know the cultural roots of Spanish society in order to explain how accepted and admitted government surveillance is.

2.1. Dictatorship government: 1939-1975
Spain was living a very unstable political and social environment during the end of 19th century and the beginning of the 20th, even taking into consideration that Spain was not involved in 1st World Ward. We had eleven months of 1st Republic with 4 presidents, a Monarchy with conservative and liberal governments, a Primo de Rivera putsch, and a 2nd Republic (1931-1936) as the main political keystones to understand this complex period of our history. On July 18th 1936 Francisco Franco lead a military putsch against the 2nd Republic, that provokes a Civil War among Spanish during 4 years (1936-1939). In April 1st of 1939, the Civil War is over and Spain began a long period of dictatorship government with Franco (1939-1975).

Franco persecuted political opponents, censured the media and otherwise exerted absolute control over the country. During this period technologies were not as developed as nowadays, but Spanish citizens were under different kinds of non e-surveillance, being a risky decision to complain about this kind of vigilance due to the lack of democratic legal guaranties.

2.2. Transition to democracy: 1975-1981

After dictator's dead in 1975, Spain move into a democratic transition stage. Franco leaves as his successor to Juan Carlos I, who starts transition to democracy. On 27th November the official coronation ceremony was held. Two days later, a Royal Pardon is promulgated and 5.655 prisoners were released. Among them, Marcelino Camacho, a union leader that in 1957 was imprisoned for his union activities. In 1976, Santiago Carrillo, leader of Spanish communist party in the exile, returns clandestinely to Spain; and it is promulgated as well the Law on freedom of assembly. These two facts can help us to understand the surveillance practices made by dictatorship government, with fatal consequences for opinions different to the "officials" ones. In 1977 was celebrated the first democratic election of current democratic period.

It is a complex period, where Spanish society is divided among the followers of Franco's legacy dictatorship government, and the supporters of an emerging democratic political system. Spanish society is under the threat of terrorist groups such as ETA and GRAPO that press for continuity or change in our political system. In 1978, Spain voted and approved the new Constitution.

In this period emerging institutions realized about the need to regulate citizen surveillance practices. As a result, in 1978, Spanish Constitution established a new democratic framework, including in its Chapter 2 (Fundamental Rights and Public Freedoms) the Right to Privacy in Section 18 [2]:

1. The right to honour, to personal and family privacy and to the own image is guaranteed.
2. The home is inviolable. No entry or search may be made without the consent of the householder or a legal warrant, except in cases of flagrante delicto.
3. Secrecy of communications is guaranteed, particularly regarding postal, telegraphic and telephonic communications, except in the event of a court order.
4. The law shall restrict the use of data processing in order to guarantee the honour and personal and family privacy of citizens and the full exercise of their rights.

In 1981 a coup attempt is suffered. The Congress of Deputies is assaulted by Lieutenant Colonel Antonio Tejero, from Civil Guard Corps, during the second vote for investiture as president of government of Leopoldo Calvo-Sotelo. Fortunately the coup did not success with the crucial intervention of Juan Carlos I, President of the Parliamentary Monarchy. The democratic system was reinforced. In 1982, PSOE (Socialist Spanish and Workers Party) reached the power in a democratic election. We can consider with this government alternation that the democracy in Spain was consolidated.

2.3. Democratic period: from 1981

Since 1981, we had several democratic elections. Democratic institutions were developed and Spain joins EEC (Economic European Community) in 1986. European culture and homogenization of legal, social and economic dimension are in a convergence process up to date. Among them, we see that the privacy concept and its legal regulation must be understood now under the European framework, taking into consideration as well specificities that Spanish society has.

The already commented Section 18 of Spanish Constitution is developed by an Organic Law. According to Spanish legal hierarchy, an Organic Law has an intermediate status between an Ordinary Law and the Constitution, and Privacy as a Right included in Chapter 2 must be regulated with this legal procedure. Organic Law 15/99 on Personal Data Protection (LOPD) states that in Spain everyone is entitled to know who, for what, when and why his/her personal data is used, and it is allowed to decide about its use. The applicable legislation regarding personal data protection is Organic Law 15/1999 and Royal Decree 1720/2007, which approves the regulation regarding data rights to access, rectify, cancel and oppose (the ARCO rights) [3].

- Access: Right to know what personal data are contained in a file.
- Rectification: Right to rectify incorrect or incomplete data in a file.
- Cancellation: Right to cancel and block incorrect data in a file.
- Opposition: Right to oppose certain, specific processing of personal data within a file.

These rights have the following characteristics:

- They are personal rights. They may only be exercised by the affected party, the legal representative of the affected party and the voluntary representative of the affected party.
- They are independent rights. It should not be considered that exercising one of these rights is a prior requisite to exercise another.
- They are free rights. Exercising this right may not incur an additional income for the file manager.

2.4. Government surveillance in Spain

Nowadays we live in a world where Internet is part of our everyday activities. In 2014, 76,2% of Spanish population has Internet access, and 74,4% of Spanish houses have Internet access. Most of Internet access is with broadband (73%) and mobile devices are becoming one of the most important Internet accessing tools. In the case of Smartphone, 81,7% refers it as the main accessing device to Internet [4].

Those facts represent an opportunity to improve economic and social welfare, but it represents as well a risk about our privacy as the *European Parliamentary Research Service* points out in their *Mass Suveillance, Risk, Opportunities and Mitigation Strategies* report [5]. This document identifies the risks of data breaches for users of publicly available Internet services such as email, social networks and cloud computing, and the possible impacts for them and the European Information Society. It presents the latest technology advances allowing the analysis of user data and their meta-data on a mass scale for surveillance reasons. Regarding the government surveillance, France approved a few days ago a law regulating national and international espionage. The rule legalizes the use of methods and "exceptional" technologies (including the use of space antennas and a tracking algorithm of communications) to control, monitor and prevent crimes and

attacks of various kinds [6]. According to European Parliament cited report [5], Government Intelligence Agencies intercept an enormous amount of information about their citizens. They use their own technological resources, hacking techniques and use of technological holes, or with direct request to technological companies that hold their user's data.

We assume that espionage between governments always had existed, and always will; and we assume that computer systems are vulnerable, but we do not assume that governments can spy their citizens without any legal control [7]. In Spain we can find the famous CESID case. CESID was the Centre for Defence Information that in 2001 become in CNI: Centro Nacional de Inteligencia (National Intelligence Centre) [8]. During 11 years, the government agency was spying and record private conversations. These illegal practices were done with politicians, diplomats, relevant business persons, journalists or even the King of Spain [9]. The scandal was discovered by press, and it had important consequences in our legal and political system. After these illegal facts, the Intelligence Spanish Centre (CSID) changes its name (to CNI) and its structure: for the first time the director was civil, not a military. Spanish jurists were blunt: listen to telephone conversations (wired or wireless) without judicial authorization deserves criminal sanction. We can see a parallelism among the analysed Spanish case and the Snowden case.

Taking into consideration emergent technologies, as José María Blanco, director of the Centre for Analysis and Forecasting of the Civil Police (Guardia Civil), in the following years we will live in a more controlled state, not just increasing the number of video cameras on public zones, even more controlled through mobile devices and Internet [10].

Within this context, we analyse the perception in Spanish students about these and other kinds of surveillance by governments and institutions. Our departure point is Snowden revelations, and we have 234 responses from 2 universities (Burgos and Rovira i Virgili). We found that students feel Internet companies (as Google, Twitter, Facebook or Yahoo!) where more invasive in their privacy than Spanish Secret Service (CNI) or other government agencies. We will show our findings in information privacy in Spain.

3. OVERVIEW OF THE SPANISH SURVEYS

This paper reflects a particular study about information surveillance by government focused in Spain and the impact Snowden´s revelation has had in youngest population. The contributions are immersed in a general research focused on analyzing cross-cultural analyses of the attitudes and social impacts of privacy, security and surveillance around the world, more specifically in Japan, Europe, New Zealand, Republic of China and Mexico. In order to achieve the purpose of the study, the authors conducted a survey during July 2014 with a total number of 3028 respondents. As this paper highlights the outcomes of Spain, the overview of the results will be focused on the 234 survey responses (University of Burgos –UBU- and University of Rovira I Virgili –URV-). Field research and sample present following characteristics:

- 234 survey responses (42% UBU students and 56% URV students) all of them valid.
- The data collected for the study is representative of the total population, as standard error estimation[1] is around 91,07%:

$$SD=[(N-n)/N]^{0,5}*(1/n)^{0,5}=0,1524 \quad (1)$$

- The survey for Spain displays 37 questions meant to collect Spanish youngsters' attitudes and behavior towards privacy and Snowden´s revelations; it has been divided in six chapters: data sample and generic field, threaten to privacy, the right to privacy, organizations and Snowden´s revelations. Some questions were asked to answer in a single- or multiple-choice form and others were requested to answer in open-ended form. Quantitative and qualitative answers display a set of information that will provide an initial insight into the privacy in ICTs that will complete the cross-cultural research with the other countries of the project.

The age spread of responders is heavily weight to young-age: 65% of the responders are (18-20) years old, 22% between (21-24) years old, and 12% of responders is older than 25 years old and the response group was 50% female and 50% male (Table 1).

A 57% of the respondents are currently studying in the field of Social Sciences, Law and Humanities. The 36% of the respondents are studying in technological and engineering careers.

Table 1. Spanish respondents attributes (%)

Gender	**Male**				**Female**			
	(50%)				(50%)			
Age	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**
	65%			22%				12%

The outcomes of the research are presented in the next section

4. SURVEY RESULTS AND DISCUSSIONS

4.1. Spanish Circumstances Related to Snowden's Revelations

4.1.1 Attitude towards the Right to Privacy in Spain

In this section, it is pointed out how relevant is the concept of right to privacy for Spanish citizens.

The concept about right to privacy has been unquestionable. Far fewer responders, a mere 1.3%, do not consider the right to privacy such an important thing. This result is alarming because where almost the total of the responders value the right to privacy very important, only a reduced percentage of them take with moderation the use of the electronic devices and read carefully the rules for privacy the companies offer to the users.

[1] Standard Error Estimation has been selected in order to understand how significant are the results considering the sample size and sampling fraction. Among others, [11] define this correlation.

The reasons that motivate why so important the right to privacy is are very diverging. To see a more detail statistical analysis of the qualitative answers, data has been regrouped. Figure 1 depicts the opinion amongst the responders. Among others, the most relevant answers have been: privacy is an important concern because suppose a life without problems, supposed to feel protected and secure, suppose a right that is within the freedom concept, and, finally, suppose "a choice". The general opinion gather through this question is mainly, that privacy means each person is free to choose with and without want to share their lives. As far as the survey results show, a 30% reported that privacy is the power to decide above our own lives. Another 30% reported privacy as synonym of freedom, and 20% picture the concept of privacy as "the space in which we can be protected against the world".

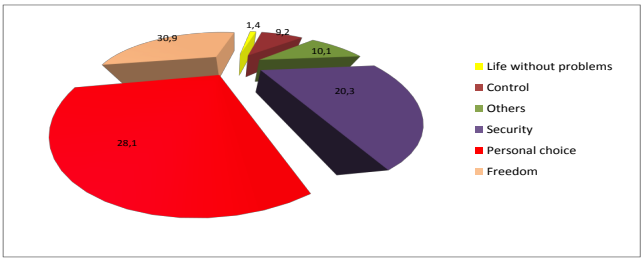


Figure 1. Why your right to privacy is important?

[12] say that "a 26.7% of the users of nets expose the data published in their profiles to strangers and, even a 4.3% of the users do not know the level of privacy of their profiles". The research carry out demonstrates that "a 16% of the users of social nets share information only with some friends; a 52% say that their information only can be seen by their friends; a 19.3% share with their friends and friends of their friends; a 7.4% share with all users and the rest of the sample they never meet before".

Almost all responders (and this is a spread certainty amongst Spanish citizens as well) claim that the use of the social nets attach a losing of privacy and this assertion is respected. Does this information mean that everyone assumes that the user that hung personal details in social nets is the responsible of all the information shared in there? This leads to the valuation of the behavior of the Spanish citizens regarding privacy and surveillance; long time ago, [13] explained this behavior in his pyramid based on the "social recognition". And [14], quoting Niedzviecki more recently, reinforced this argument saying that "people want to reveal, they want to be known, they want to be seen". Therefore, is this necessity what is causing the overexposure of the people in the social nets? Is this necessity over the fear to be threatening knowing certainty that all our comments and pictures will be shared from one side to the other of the world? On the other hand, where are the limits of the law? What means the right to privacy?

In Spain, The Data Protection Law is currently the reference of the jurisprudence. Faraway of this law, Europe does not have right now any other regulation or law to control this lacks of privacy or intimacy. In this space where the law is not applicable appear the ethics and the Deontology. Both, jurisprudence and ethics make up the scene where currently Digital Natives converge. Trying to clarify this applicability, [15] -who is in charge of the Central Department of Crimes in ICT of the Ertzainzta Police in Spain- analyses the barriers that involve the investigation of Crime in ICTs and argue two scenarios: the spatial scenario and the temporary scenario.

Regarding the first one, the spatial scenario, there is a cooperative mechanism such as the National Centre of INTERPOL to help in the investigations. But the requirements to receive support are two: (i) the quantity cheated must be over a number and (ii) the crime done must be proved that has been carried out by an organized net. In this sense, investigations turn dark and complex as it is really touchy to confirm that the crime is perpetrated by an organized net indeed, and, on the other hand, the quantities cheated sometimes are not over the limit to be a crime, therefore, the crime is unpunished. Regrettably, there is a lack of consensus in the legislation. Law is territorial and what is a crime in a state is not in other. 23rd of November of 2001 in Budapest, the Europe Board draft the Cybercrime Agreement with legislation for 49 countries involved in the project. In 2007 only 8 out of 49 countries had ratified. As control measure it was developed the Framework Resolution 2005/222/JAI of the Board 24/02/2005 which legislation is focused on the ICT attacks.

As far as the temporary scenario concerns, the intelligence support services cannot cover all cybercrime investigations because the criminals find weakness in the ICTs that make easy for them to carry out crimes. Say, among others: the deadline of the logs (ISPs[2]), the risk of destruction of evidences in the case of the victim (LSSi and CE stablish a maximum time for keep the logs during one year but nothing about minimum time of keeping) and the risk of destruction of evidences in the case of the perpetrator".

It is remarkable the tiny portion (2.25%) of responders that recognize they do not understand what the right to privacy means; in contrast, 84% say that they understand perfectly well and 12% understand but with some difficulties. This is partially correct and odd because considering making the practice of ask anyone under which law of privacy is working Facebook, no one knows about it. Same circumstance happens if we ask about the declaration of rights and responsibilities of Facebook, about the share contents or information, about security, about rights and protection of others, amendment, the conflicts, etc. One thing is the idea about what means security and privacy, and, other thing is regulations, rights and responsibilities we have as users.

The meaning of right to privacy has been understood from different perspectives. Figure 2 shows the group of qualitative answers. Only a minority of responders (16%) considered that the right to privacy is to make what you want without being spying; few (34%) recognized that the right to privacy is to feel secure and with your private life protected; a larger group of responders (54%) considered the right to privacy is the right to decide who is the person or persons you want to share your private life with and 27% said that it is the right to be respected without meddling. As far as the survey results show, while all the responders seemed to take into account the relevance of the right to privacy, only a minority take measures. Quoting the research about cybersecurity, carried out by [12] "a 42% of the users do not use active security measures; a 69,4% of the users say that the updating of security is done automatically in the Pc's and the 57.8% of the users never check the electronic devices to look for virus letting the antivirus program do by itself. Surprisingly, there is a 12.4% of the users that never protect the Wi-Fi net".

[2] In Spain, in 2007 was impossible to make a request form in an ISP about the identity of a person with an IP if the request was in a period of time bigger than one year.

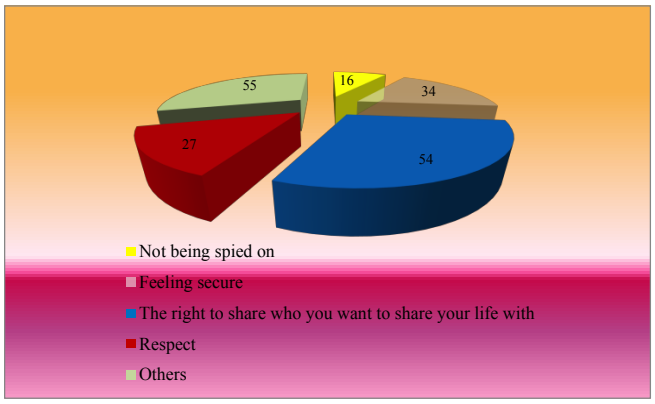


Figure 2. What means the right to privacy?

Taking into account these outcomes, questions of the survey regarding threaten to privacy such as risk recognition in the Internet and non-Internet activities, provided interesting findings about youngsters´ feelings. Figure 3 shows the difference in risk perception of privacy invasion between the two types of activities.

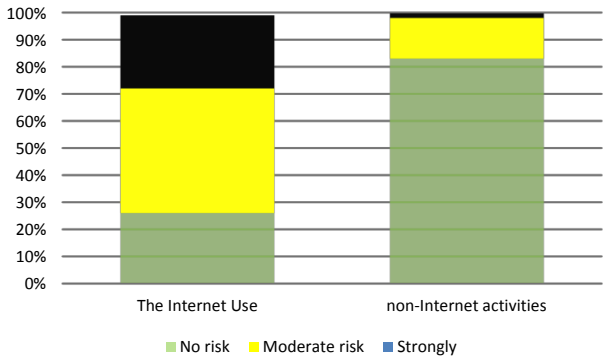


Figure 3. Risk recognition in the Internet and non-Internet activities

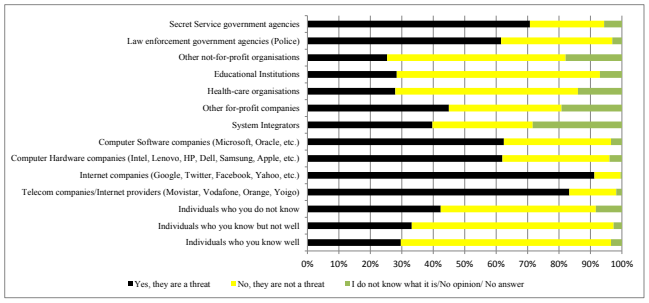
A 27% of the sample feels vulnerable when using Internet due to privacy threat. A 46% believe that there is a moderate risk and a 26%, nevertheless, consider that using Internet does not entail a risk for the privacy. The reduced percentage of respondents aware of the problems that internet can bring to them do indicate that a sizeable proportion of young people do not have knowledge about the wide range of cybercrimes and the consequences of a wrong use of the net.

Furthermore, the risks that entail the non-use of Internet changes dramatically. All in all, we assert that there is no certainty knowledge about the threaten derived from activities no linked with Internet, because, only a 1.7% of the respondents believe that all the tasks carried out off the net are not susceptible to be hacked, copied, etc. On the other side, a 15% of the sample has valued this risk with medium level of importance. Finally, there is high percentage, the 83% that consider there is no evidence of risk linked to it. This result is very meaningful. It is remarkable the high percentage of users that are not aware of being threaten by the cyberattacks when no using the net. This lack of awareness reveals the level of knowledge about this issue. [16], point out in his book about the psychology of cybercrime that there are three types: crimes carry out off the net but spread thank to internet, crimes that did not exist in the past before internet arrival and crimes carry out by people using their online avatars.

In order to understand Spanish youngsters' perception about sources of privacy invasion, we asked a threat level from organization to and from technology to technology. Table 2 shows the average scores of each group as a source of privacy invasion (maximum score is 1 and min score is 4). Surprisingly, top three groups viewed as a threat of invasion were "Internet companies (M=1.46)", "Telecom companies (M=1.78)" and "Secret Service Governments (M=1.9)". On the other hand, regarding the average scores of the technologies (Table 3), respondents indicate that "Smartphones (M=1.58)", "Payments online (M=1.72)" and "Online Shopping (M=1.88)" are ranked in the top of threaten technologies.

Table 2. Ranking of groups that are viewed as a threaten to privacy

Q8. How much do you feel that the following groups threaten your privacy?		
Groups	**Means**	**S.D.**
Individuals who you know well	2,97	1,037
Other not-for-profit organisations	2,91	.823
Educational institutions	2,84	.773
Health-care organisations	2,81	.869
Individuals who you know but not well	2,8	.765
Individuals who you don't know	2,62	.996
System Integrators	2,4	.963
Other for-profit companies	2,35	.909
Other government agencies (Health, Interior, Tax, etc.)	2,33	.945
Computer hardware companies (Intel, Lenovo, HP, Dell, Samsung, Apple, etc.)	2,2	.905
Computer software companies (Microsoft, Oracle, etc.)	2,18	.930
Law enforcement government agencies (Police)	2,16	1,125
Secret service government agencies (CNI)	1,9	1,002
Telecom companies/ Internet providers (Movistar, Vodafone, Yoigo, Orange)	1,78	.840
Internet companies (Google, Twitter, Facebook, Yahoo!, etc.)	1,46	.723



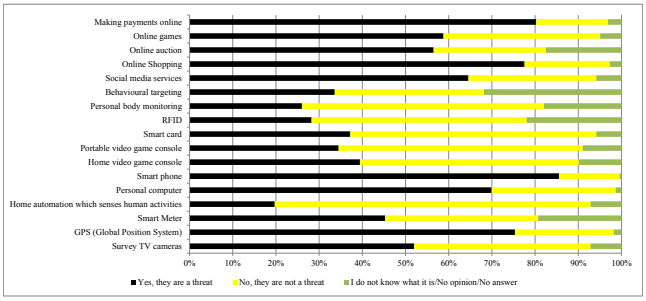
Users consider that the organizations that can damage more our privacy are, in a 90% the social nets and with an 80% the telecom companies. Additionally, police and the secret services are also a threat as users consider they spy citizens.

Nevertheless, is very striking that users do not consider their friends or relatives a threat; this is the way of thinking of the 70% of the respondents; regarding this percentage, is interesting to point out that the data of some research done by United Press International (2008) in [16] state that more than the 40% of the teenagers of USA have been victims of cyberbullying and only one out of ten have told off their parents. As well as the research of Microsoft (2009) saying that almost third part of the European teenagers had been cyberbullying victims. And this bullying can come from people that know you well, a little or just nothing. On

the other side and going back to the results of the survey, it is important to show the lacks of knowledge and information about the System Integrator companies; only a 30% of the respondents say that they do not know the meaning or motto of these companies. Also, there is a certain lack of knowledge about the activities or objectives pursuit by for-profit companies and not-for-profit companies.

Table 3. Ranking of technologies that are viewed as a threaten to privacy

Q9. How much do you feel that the following technologies threaten your privacy?		
Technologies	**Means**	**S.D.**
Home automation which senses human activities (e.g., air conditioner, lighting apparatus)	3,12	.862
Personal body monitoring (Fitbit, etc.)	2,88	.888
Portable video game console (PSP, Wii-U, etc.)	2,74	.892
RFID (Radio Frequency Identification)	2,73	.827
Smart card (transport card, gym card, etc.)	2,72	.907
Home video game console (Wii, PlayStation, XBOX, etc.)	2,64	.889
Behavioural targeting	2,47	.898
Survey TV cameras	2,39	.890
Smart meter (an electricity meter providing your supplier with regular, approx. every 30 minutes, readings of your usage)	2,34	.885
Online games	2,27	.849
Online auction	2,17	.861
Social media services	2,08	.953
Personal computer (Widows machines, Mac, etc.)	2,04	.936
GPS (Global Positioning System)	1,94	.886
Online shopping (Business to Consumer ecommerce)	1,88	.813
Making payments online	1,72	.835
Smart phone (iPhone, Android, etc)	1,58	.762



A majority of responders (85%) said that in their opinion, the first damage comes from the Smart Phones; the second comes from online payments and the third and fourth position is for the online shopping and GPS. As noted above, there is an illogical behavior between thoughts (threat in ICTs) and decision-making (acquisition of electronic devices). On the one hand, a significant majority perceive there is a threat in the technologies, and, on the other hand there is a relax attitude to it. It is very odd that even being aware in a big percentage of threat with the electronic devices such as the Smart Phones, the amount of sales of these products is increasing substantially. Quoting [12], "a 86.8% of the cyberusers with a high frequency of connection to the nets held a Smartphone or a similar electronic device such as an intelligent handy"

4.1.2. The Degree of Recognition of and Interest in Snowden's Revelations in Spain

Although the concept of WikiLeaks is a new one in Spain, display of information about Edward Snowden has been more spread than Assange´s. Results of the survey prove this hypothesis. 60% of responders have heard about him. Surprisingly, the percentage of people that has not heard about him continue being relatively high (38%).

Amongst those who claimed to know Snowden and Assange´s revelations, some of them had downloaded this information from the TV programs (64%), online (21%) or printed newspapers and the social nets (10%). Chatting with friends is also a source for sharing information and, finally, the lectures at the university are the place where they less have heard about it (5%). As far as the results show in the survey would be interesting to analyze and focus our attention on the high percentage of responders that have revealed they have been updated about this information in the TV, in comparison with the students that have read online news. Considering this result we can provide the research with a new target questions such as:

- Does it mean that the youngest spend more time watching TV than surfacing nets?
- Which is the purpose of the users of electronic devices?
- Are they more a device for play games, chat with friends, participate in the social nets…instead of using also for learning from the platforms?

At the time of the survey, around a half of responders have spoken with others about Snowden WikiLeaks (51%); this percentage is interesting because it can be concluded that the students seemed to have curiosity about who was this person and why Snowden was trending topic in the nets. Surprisingly, amongst this percentage, 46% has looked for more deeply information about Snowden revelations and 43% of them claim that have got a lot of information about the situation

As reported above -in the question 8-, amongst the percentage of responders (65%) that considered the social nets as a threat for their privacy, the ratio of the sample that protect their systems with new passwords should be high; the survey results validate this hypothesis. 39% of responders continue without protecting their electronic devices while 52% of responders has taken measures to be more protected, such as: "pay attention to all the information published on the social nets", "change the characteristics of privacy of some electronic devices", "reduce the use of some electronic devices or erase some personal data and content of the social nets".

Considering the responses to this section, almost half Spanish responders seem to have an awareness of the importance of privacy and security and the consequences of being unprotected to a certain extent.

4.1.3. Evaluation of Snowden's Activities in Spain

Spanish respondents judge the activity of Mr. Snowden spreading some opinions. Figure 4 is a picture with all the qualitative responses; it depicts the reasons why Mr. Snowden revealed the secret information, made off some documents and whistle-blowing. Quarter of responders (26%) have considered that the reason why Mr. Snowden told off was because during many years at his post in CIA, he had discovered unjust, odd and illegal

movements in the haul of documents. And he had the moral obligation of fighting for stand up for the right to privacy any citizen deserves. Another minority, 17% of responders, think that he had the necessity of spread the world to be aware of privacy and security because we are being spying continuously and the mandatory obligation of taking measures to avoid this kind of crimes (we are not as secure as we think we are and anyone can follow our steps just clicking in a bottom). But in general and general speaking, it is impossible to achieve this level of privacy, even staying off the nets; in fact, a secure way of get privacy, face to face with so developed technology of big institutions such as CIA or Pentagon, would be not running through internet, neither any kind of telephone, no bank transfers nor online shopping, etc.; all in all, go back decades and live as an isolated person.

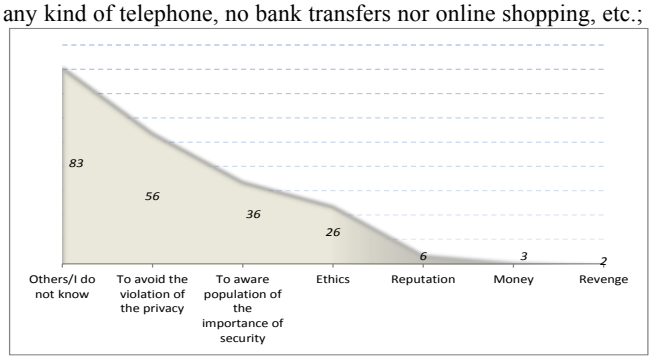


Figure 4. Why do you think Snowden determined to make those revelations?

This is an interesting assessment as we can evaluate if the benefits reached thank to ICTs are not against of the disadvantages of being so evolved. A second debate arises with the question: is on us the security of the electronic devices or has nothing to do with us because is only on the Intelligence Agencies?

Moreover, almost 70% of the responders agree and consider that what Mr. Snowden did served for many purposes. However, 16% of responders hardly consider that it served any purpose and similar rates, (16%) of responders, prefer not to answer. These percentages are even higher than results obtained in a similar pilot survey conducted in Spain by the authors in June 2014 (46%).

4.2. Empirical Consideration about Influence of Snowden's Revelations

After evaluating the general results of the survey it is mandatory to cross some more specific parameters in order to stablish the hypothesis and their final statements. Although all information contains in the survey is useful, for the statistical analyses of crossing ratios we have selected the questions focused on behavior (1 and 4) and Snowden´s revelations (2 and 3). Discussion on this section will be focused mainly on these four statements: 1) Does the concept of privacy and security have any influence on the users that are more aware of the risk involved in the use of the Internet?; 2) Users who rely on the good effects Snowden Revelations have had, are more compromised with doing the same?; 3) Users more updated about WikiLeaks effects are more aware about the risk of privacy and security? and 4) Do the respondents who have selected social nets as the group which more threaten their privacy are more committed with changing the way of communicating?

4.2.1 Does the concept of right to privacy and security have any influence on the users that are more aware of the risk involved in the use of the Internet?

Methodology followed in the procedure for examining this research question was done making a division between the respondents they considered the use of Internet a risk or not (Q6) (Do you feel that your use of the internet involves taking risks with your privacy?) Then one concrete question was empirically considered via T-test in which above these two groups were used as control and treat group. The question is whether the respondents who felt that the use of Internet involved risk with their privacy tend to validate privacy as important compared to those who did not consider the use of Internet as a threat. The result of the T-test (Table 4) reveals that the mean of the group who feels the use of Internet as a threaten for privacy (M=1.27, SE=.038) does not exceeds that of the group who doesn't feel (M=1.33, SE=.063) and the difference between these averages (DM=.063, 95% CI [.012, .115]) is a statistically significant at one percent significance level ($t(185) = 1.97$, $p < .01$). That is, we were statistically able to confirm that the respondent group who feel the use of Internet is a risk for privacy does not tend to consider the right to privacy more than the group who don't feel that.

Table 4. Results of T-test: Q6 vs Q10

Means of each group	Answer to Q6	Yes	No
	Mean	1.27	1.33
	St. Error Mean	.038	.063
Difference between Means	Mean Difference	.063	
	95% CI	.012 ~ .115	
Statistics	t-value	1.97	
	d.f.	185	
	p-value	< .01	

4.2.2 Users who rely on the good effects Snowden Revelations did are more compromise with doing the same?

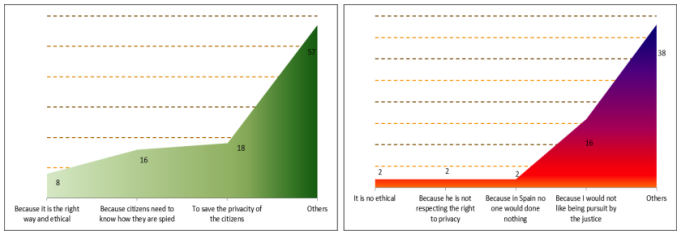
The second statement is whether there is the tendency that respondents who rely on the good effects Snowden Revelations had in society would be more compromised to do the same in comparison with the group against this kind of behaviour. To confirm it average scores of Q28 (Have Snowden's revelations served the public interest or harmed it?) and Q30 (If you were an American citizen and were faced with a similar situation to Snowden, do you think you would do what he did?) are compared between the two groups (Table 5). According to the result of T-test, the respondents who think Snowdens´ revelations served for good purpose tend to feel smaller compromise with doing the same he did (M=1.37, SE=.046) compared to those who think what Snowden did harmed more than benefit in society (M=1.56, SE=.101), and the difference between the averages (D=.188, 95% CI [.070, .307]) is significant at one percent significance level ($t(136) = 1.97$, $p < .01$). These results reflect that with independence of within the Snowden revelations harmed or benefited, Spanish respondents would do what he did. This result is coincident with a previous research done one year before [7].

It is interesting to point out that the reasons given in the survey to the questions if they were face to a similar situation Snowden was, would do the same, Spanish respondents have said, among others reasons: to safe privacy of citizens (18%), because citizens need

to know more about the spy actions they are under to (16%), because it the right way to do the things and it is ethical (8%) and 57% of responders do not know or prefer not to answer. Figure 5 shows the qualitative answers in several groups.

Table 5. Results of T-test: Q28 vs Q30

Means of each group	Answer to Q28	Yes	No
	Mean	1.37	1.56
	St. Error Mean	.046	.101
Difference between Means	Mean Difference	.188	
	95% CI	.070 ~ .307	
Statistics	t-value	1.97	
	d.f.	136	
	p-value	< .01	



Reasons that motivate what Snowden did (if I were Spanish citizen)

Reasons that motivate NOT to do what Snowden did (if I were Spanish citizen)

Figure 5. Why would you do or not, as a Spanish citizen, what Snowden did if you were faced with a similar situation to him in Spain?

4.2.3 Users more update about WikiLeaks effects are more aware about the risk of privacy and security?

The third statement was attempted to confirm whether the information spread in media about WikiLeaks (Q19) (Have you heard about Snowden's revelations?) served the population to be more aware about the implications privacy and security have got (Q6) (Do you feel that your use of the internet involves taking risks with your privacy?). The result of the T-test (Table 6) reveals that the mean of the group who had heard about Snowden revelations do not tend to be more aware about the risk of privacy and security (M=1.97, SE=.072) that of the group who did not heard about Snowden revelations (M=2.09, SE=.080) and the difference between these averages (DM=.125, 95% CI [.108, .142]) is a statistically significant at one percent significance level (t (213) = 1.97, p < .01). That is, we were statistically able to confirm that the respondent group who had heard about Snowden revelations is not more aware of privacy and security than the group who did not heard about him.

Table 6. Results of T-test: Q19 vs Q6

Means of each group	Answer to Q19	Yes	No
	Mean	1.97	2.09
	St. Error Mean	.072	.080
Difference between Means	Mean Difference	.125	
	95% CI	.108 ~ .142	
Statistics	t-value	1.97	
	d.f.	213	
	p-value	< .01	

4.2.4 Do the respondents who have selected social nets as the groups which more threaten their privacy are more committed with changing the way of communicating?

The fourth statement was attempted to confirm whether the group of respondents that had chosen Social Nets as the more dangerous group that threaten their privacy (Q8-e) (How much do you feel Social Nets threaten your privacy?) had got a higher disposal for changing the way of communicating online (Q24) (Have you changed your way of communicating online since you heard about Snowden´s revelations?). The result of the T-test (Table 7) reveals that the mean of the group who had selected social nets as the groups which more threaten their privacy do not tend to make any changes in their way of communication (M=1.73, SE=.034) in comparison to the group who did not consider social nets as a threaten (M=2.09, SE=.0125) and the difference between these averages (DM=.015, 95% CI [-0.288, 0.188]) is a statistically significant at one percent significance level (t (182) = 0.122). Statistically we can point out that Spanish citizens are aware of the risk social nets have regarding privacy, but, do not take measures in order to avoid negative consequences neither higher damages.

Table 7. Results of T-test: Q8-E vs Q24

4.2.5 Do we approach Snowden´s Revelation depending on our gender?

Male are more updated about Snowden´s Revelations in comparison to females. Taking that into account it has been consider de possibility of analyse if the estimation of the impact of the revelations is gender dependent. Thus Q28 (Have Snowden´s Revelations served the public interest or harmed it?) has been split depending on the gender of the subjects. As shown in Table 8, male and female are both positive about the impact that revelations had in society although female do show a higher lack of knowledge and interested in the revelations compared to their male colleagues. Based on these result, it seems to be thought that gender is not driven factor in the approach to this question.

Table 8. Statistic Results: Q1 vs Q28

Means of each group	Answer to Q1	Male	Female
	Mean	1.85	1.96
	St. Error Mean	0.083	0.102
Difference between Means	Mean Difference	.103	
	95% CI	.066~ .141	

5. CONCLUSION

Snowden´s Revelations have been for Spanish citizens the confirmation of something that was in their minds. There is a common feeling of being spied, followed or even controlled while using internet [17]. Survey outcomes demonstrate that a majority of respondents are aware of Snowden´s Revelations, although only a few consider seriously taking actions to improve their privacy in internet. Truthfully, media in Spain has refrained from

talking about Snowden´s update details and as consequence that is damaging the interest of our youngest about the topic.

One of the most relevant findings is that Spanish citizens are committed with losing privacy in benefit of the society and they feel the necessity of spread the world to be aware and try to stop this kind of crimes. The big doubt is how to deal with this situation, as they consider there is impossible to achieve any level of privacy, even staying off the nets; in fact, a secure way of getting privacy, would be not running through internet, neither any kind of electronic devices. It means to go back decades and live as an isolated society. All in all, would be really interesting to evaluate if the benefits reached thank to the TIC´s are not against of the disadvantages of being so evolved.

6. ACKNOWLEDGMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006 and the JSPS Grant-in-Aid for Scientific Research (B) 25285124 and (B) 24330127.

7. REFERENCES

[1] García de Cortazar, F.; Gonzalez J.M. 2012. *Breve Historia de España.* Alianza Editorial. Madrid.

[2] Constitución española. 2015. BOE núm. 311, 29 de diciembre de 1978. http://www.congreso.es/consti/index.htm. (Accessed 05/01/2015).

[3] Government of the Principality of Asturias (2015): ARCO Rights, https://sede.asturias.es/portal/site/Asturias/menuitem.fe57bf7c5fd38046e44f5310bb30a0a0/?vgnextoid=290c7a0266719210VgnVCM10000097030a0aRCRD&i18n.http.lang=en#question (accessed 01-15-2015).

[4] INE. 2014. Notas de presa: Encuesta sobre Equipamiento y uso de Tecnologías de Información y Comunicación en los Hogares 2014; http://www.ine.es/prensa/np864.pdf (accessed 01-15-2015).

[5] European Parliamentary Research Service. 2015. Mass Suveillance, Risk, Opportunities and Mitigation Strategies report. http://www.europarl.europa.eu/RegData/etudes/STUD/2015/527409/EPRS_STU%282015%29527409_REV1_EN.pdf (accessed 01/15/2015).

[6] Arrieta E. 2015. ¿Nos espían los gobiernos?. Diario Expansión, 05/23/2015. http://www.expansion.com/tecnologia/2015/05/23/555cc862ca474168478b45b8.html (accessed 05/23/2015).

[7] Delle Femmine, L. (2014). El gobierno español hace muy buen espionaje, entrevista a Eugene Kaspersky; http://tecnologia.elpais.com/tecnologia/2014/05/05/actualidad/1399281568_671465.html (accessed 01-15-2015).

[8] CNI (2015): El CNI, al servicio de España y de los españoles; Centro Nacional de Inteligencia; http://www.cni.es/es/queescni/historia/elcni/ (accessed 01-15-2015).

[9] Galiacho, J.L. 2007. Las escuchas del CESID. Los espías del gobierno grababan hasta al rey, 18 historias que cambiaron España, El Mundo; http://www.elmundo.es/especiales/2007/10/comunicacion/18elmundo/cesid.html (accessed 01-15-2015).

[10] Ballesteros R. 2013. El Centro de Prospectiva de la Guardia Civil pronostica una sociedad vídeo vigilada antes de 2030, lainformacion.com; http://noticias.lainformacion.com/espana/el-centro-de-prospectiva-de-la-guardia-civil-pronostica-una-sociedad-video-vigilada-antes-de-2030_WjlPAkMoqAYkoHoNbfb387/ (accessed 01-15-2015)

[11] Lininger, Ch. and Warwick, D. 1985: *La encuesta por muestreo: teoría y práctica*. CECSA. México.

[12] INTECO (Instituto Nacional de Tecnologías de Comunicación) (Gómez, M. and others) 2014:"Estudio sobre la ciberseguridad y confianza de los hogares españoles".

[13] Maslow, A. H. 1954: *Motivación y Personalidad*. Sagitario.

[14] Lamb, G. M. 2009: "How we are losing our privacy online". CS Monitor. com

[15] Viota, M. and others 2007: "Problemas relacionados con la investigación de los denominados delitos informáticos (ámbito espacial y temporal, participación criminal y otros)" in *Cuadernos penales Jose María Lidón, nº 4. Delito e Informática*. *Algunos aspectos*. Universidad de Deusto.

[16] Kirwan, G. and Power, A. 2012: *The psychology of cybercrime: Concepts and Principles*. Information Science Reference. IGI Global. EEUU

[17] Murata, K., Adams, A. A., Orito, Y., Fukuta, Y. and Lara Palma, A. M. (2014): "Social Impacts of Snowden's Revelations in Japan: An Exploratory Research". Proccedings of the 4th International Conference on Easy Privacy: Asian Privacy Schoolars Network (APSN 2014).

Surveillance of information and personal data by Mexican government: The social impact in Mexican Citizens

Juan Carlos Yáñez-Luna
Universidad Autónoma de San Luis Potosí
Av. Pintores S/N, 78213, San Luis Potosí, México
+524448131238 ext. 112
jcyl@uaslp.mx

Mario Arias-Oliva
Universitat Rovira i Virgili
Av. Universitat 1, 43204 Reus, Spain
+34977759800
mario.arias@urv.es

Kiyoshi Murata
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2165
kmurata@meiji.ac.jp

Pedro I. González Ramírez
Universidad Autónoma de San Luis Potosí
Av. Pintores S/N, 78213, San Luis Potosí, México
+524448131238
pedro.gonzalez@uaslp.mx

Andrew A. Adams
Meiji University
1-1 Kanda Surugadai
Chiyoda, Tokyo 101-8301, Japan
+81 3 3296 2329
aaa@meiji.ac.jp

Ana María Lara Palma
University of Burgos
Avda. Cantabria, s/n
09006 Burgos, Spain
+34 947 259 360
amlara@ubu.es

Andrew A. Adams

ABSTRACT

This study analyses the perceptions about Edward Snowden's revelations in Mexico. A questionnaire survey was developed and applied to students in a Mexican University (Autonomous University of San Luis Potosí). This Study is part of a global research about privacy perceptions by young people in different countries.

Categories and Subject Descriptors

K.4.1 [**Computers and Society**]: Public Policy Issues – abuse and crime involving computers, privacy, use/abuse of power

General Terms

Security, Human Factors, Legal Aspects.

Keywords

Surveillance, Privacy, Mexico, Rights.

1. INTRODUCTION

Today, issues related to privacy and personal data protection have a major impact on Mexican society. Taking into consideration that it is very difficult to guaranty total privacy about citizen's private information even in an off line world, with a growing dependence on, Internet and information technologies increase exponentially threats in this area.

This may be due to the digitization of personal data stored in the various government or sites using web services like social networking, video on demand, electronic payments, etc.

After Snowden's revelations of the possible conspiracy of surveillance (EE.UU. and GCHQ[1]), many political and social discussions were generated in several countries. In Mexico this situation did not have an important impact on the citizens' threats. In this respect, citizens in Mexico don't have a clear idea about how this situation could affect their privacy , so people adopts a passive status [18]. Anyway, Mexican government did a diplomatic protest when NSA spied some activities of the ex-president Calderón in 2012 and some electoral process and electoral candidatures[5].

In 2014 Mexico faced a reform in the national telecom law. This reform proposes some changes in secondary laws in which were affected several topics, including Internet freedom and expression freedom of citizens, and this caused conflicts in some politic parties, provoking social movements and protests against these reforms. Legal regulations about the right to privacy in Mexico are in an early stage. People are not informed about that topics and government could take advantage of it obtaining critical private information. Because of this reason, we consider that it is important to recognize privacy rights as a priority for Mexican citizens in order to protect themselves from legal and illegal organizations [15].

This article aims to address issues of privacy and security from a social perspective in Mexico as results of Snowden's

Conference'10, Month 1–2, 2010, City, State, Country.

[1] GCHQ refers to Government Communications Headquarters, one of the British Intelligence Agencies.

revelations. To reach this goal we analyze specific information of students of the Autonomous University of San Luis Potosi were collected by a survey instrument.

The structure of the article is as follow. After this introduction we will show an overview about the politic system in Mexico and the implications in the recent events in the topic of surveillance. In the third section we will analyze and discuss the outcomes of the survey applied to Mexican students. Finally we will show our final conclusions.

2. BACKGROUND

2.1 Historical overview on Mexican politics

In pre-Hispanic era Mexico had a culture and politics based on religious beliefs and monarchical system of government in which was reigned by emperors. But the conquest of Spain in 1521 (specifically in Tenochtitlan and consequently to other indigenous populations) marked a new stage of government.

In this new period, the monarchy system changed from the imperialist regime towards a regime of Viceroyalty. However, although this type of system of government did not have a good administration, this causes an independence movement in 1810. In the period of 1920-1821 the political system in Mexico experienced a period of transition, a regency period was formed to work as executive branch. However, political movements influenced to return the monarchy system proclaiming to Agustin de Iturbide as Emperor of Mexico. Political movements in the country continued and the imperialist regime was overthrown. In 1824 the First Constitution of the United Mexican States was proclaimed, the first federal election for President and Vice President were held.

From that period until 1834 Mexico had many internal wars between liberal political movements (who considered a democratic republic) and Conservatives (who considered a European monarchy). During the nineteenth century Mexico has participated in wars to defend the country from foreign invasion that have influenced the political, social and economic system of the country. Mexico has experienced two political dictatorships in the late nineteenth century and early twentieth century in the democratic system. Adjustments to the Constitution of the United Mexican States and the establishment of an institutional democracy causing that some political parties emerged.

2.2 Surveillance and Privacy right in the recent events

Throughout the modern history of Mexico take place various security-related events that have had political and social repercussions in the nation. Such are the cases of student protest movements occurred in 1968, 1971, 2012 and cases like Ayotzinapa and Tlatlaya in 2014. Other issues are related to drug illegal commerce and corruption of public representatives, murders and kidnappings of journalists, conflict of interest between political parties, etc.

The use of information technology as a tool for information spread through social networks, emails, blogs and videos, etc.; has been fundamental. However, the use of these new technologies has many negative consequences, as in privacy and security areas. It is more often to read news about cyber-attacks, fraud or spread of malicious software [14].In Mexican case, according with Symantec [19] cyber-attacks have increased significantly to 113% in 2013 compared with the previous year.

Mexico has had several agencies involved in national security. These agencies have conducted several operational intelligence and protection of information at government level since the early twentieth century. However the society was not so involved in these events [1,11]. From the movement "I am 132" in 2012, the society began to integrate ICT in political activities. We consider that citizens began a new social and political culture, where the use of social networks to express ideas, share and disseminate information is crucial. Government surveillance was done in this new communication channels without any legal support. Mexican population does not trust very much in their public agencies. There is a doubt about the effectiveness of these agencies to clarify cases of violation of freedom expression. For example, there are cases of attacks on Mexican journalists [3] or the student protests [16]. The most known is Ayotzinapa case [6]. 43 students were murdered when they attempt to start a protest meeting. There are other cases of deprivation of expression, surveillance [10] and data protection [17].

In this respect, the Constitution of the United States of Mexico, 1917, Article 16, considers: “No one may be subjected to interference with his or her person or his or her family, to arrest, detention or imprisonment or to have his or her home searched, except in accordance with a written order from the competent legal authority, in due form and for reasons previously defined by law.”. At the same constitutional article states that:

“Private communications are inviolable. Criminally punishable by law any act that violates the freedom and privacy in communications, except when they are provided voluntarily by any of the individuals involved in them. […]. Under no circumstances any communications that violate the confidentiality established by law won’t be accepted.” [4].

In terms of expression of ideas and protection of personal data, Article 6 of the Constitution Paragraph A - Section II states that: "The information relates to privacy and personal data will be protected under the terms and subject to the exceptions set in the laws". Moreover actual telecommunications law in the Article 145 Section II and III states the following statements:

- “No discrimination. Dealers and authorized to hire the Internet providing access shall not obstruct, interfere, inspect, filter or discriminating content, applications or services.”
- “Privacy. Dealers must preserve the user privacy and network security.”

Therefore, it is important to know the perceptions that Mexican society has in relation to the issues of surveillance and protection of personal data by government institutions, private organizations and telecommunications service providers.

3. OVERVIEW OF THE MEXICAN SURVEYS

3.1 Methodology

To analyze the data in this study, specific information of students of the Autonomous University of San Luis Potosi were collected by a survey instrument. The survey was designed to obtain information about several topics such as, privacy rights perceptions of Mexican people, threat to privacy, about Manning and Snowden cases of surveillance and how technology affect their privacy. Most of the items in the survey were measured using a Likert scale (1. Strongly, 2. To an extent, 3. Not much, 4. Not at all, 5. Prefer not to answer). The survey was developed

online by the Survey Monkey service. However, we decided not to send the survey via email to students, instead of that, we applied the survey to students during a classroom session. A total of 163 surveys were answered and 2 of them were rejected by inconsistency in the data, so eventually left 161 assessment surveys.

3.2 Survey analysis.

For the first section of the survey we introduce some demographic data to know how our sample is distributed. The descriptive analysis for this section shows that the range of age is between 18 and 25 years, all of them were students with an average of 18 years old as is shown in Table 1.

Gender	**Male**				**Female**			
	(45%)				(55%)			
Age	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25+**
	89%			10%				1%

Table 1.Mexican Respondent attributes (number of respondents (%))

We also identified that most of the respondents were women with 55% against 45% were men. According to the survey, the most representative career was Social Sciences with 89% of the respondents and the 11% are divided in careers like Engineering 1%, Humanities 2%, Natural Sciences 1% and other 7%. For the nationality question, most of the respondents were Mexicans with the 98% and 2% are foreigners.

4. SURVEY RESULTS AND DISCUSSIONS

4.1 Mexican Circumstances Related to Snowden's Revelations

4.1.1 Attitude towards the Right to Privacy in Mexico

In this research, students were questioned how they perceive the importance of the Right to Privacy. The outcomes suggest that 98% of respondents consider that it is important as is showed in figure 1.

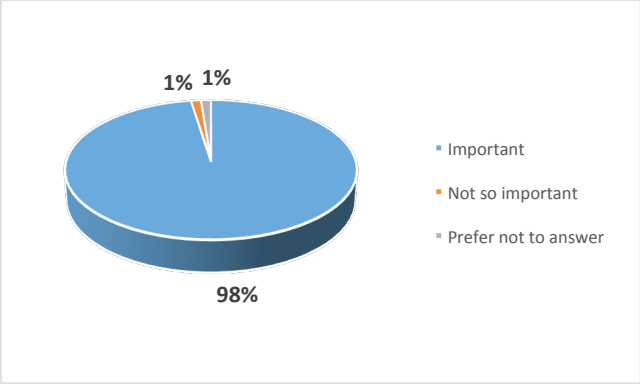


Figure 1. Is your privacy important?

However, when students were questioned about if they understand what right to privacy is, only 27% understand very well and 61% have a moderate understand about the concept as is shown in the figure 2. It is logical to assume that to have a real privacy is a utopic context, but the reality is that most of the people have a general idea of what privacy is. According to [12] the right to privacy can be defined as "the right of the individual to determine for themselves when, how and to what extent they will release personal information about themselves". Our findings also suggest that students known that every Mexican citizen have the right to data protection and privacy [4], but most of the respondents may have a partial ignorance about privacy.

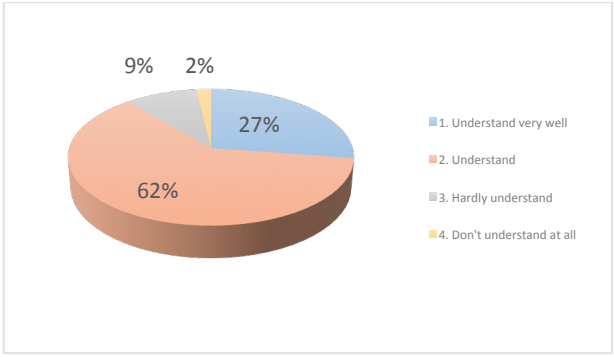


Figure 2. Percentage of understanding about Privacy's concept

In this respect, the ignorance must be due by two principal reasons. The first one is that most of the governmental organizations are perceived as threating entities and untrustworthy, and the second one is that these organizations have many subsection or subdivision and it is possible that citizens don't know about all of them and their functions. In this case, we asked Mexican students how much do they know about some governmental organizations. We decide to separate the international and national governmental organizations in order to differentiate and to obtain a better evaluation. Our findings show that 59% of Mexican students knows well international agencies like the FBI and 52% the CIA, but don't have enough knowledge about NSA: 57% of the respondents pointed that they have heard about it but don't know about their functions. The GCHQ is the least known agency, only 5% of Mexican students know about it. The figure 3 shows the outcomes of the survey.

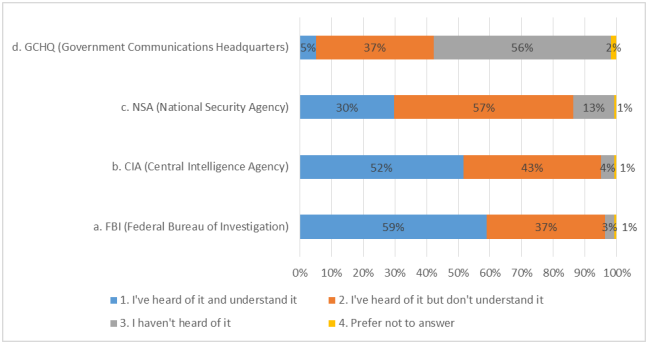


Figure 3. International Security Organizations

On the other hand, national agencies doesn't seem as the best known by Mexican people. The most well-known security agency is Policía Federal (PF) with 68% of the respondents, another one is SEDENA (Military Corporation) with 57% of the respondents. Since 2006 Mexican government launched an anti-drug strategy. Its main goal was to dismantle criminal organizations using military forces (SEDENA) and the Federal Police (PF).We can assume that through these events young Mexicans know these corporations, but also have noted that some respondents do not know the specific functions of these organizations. It will be necessary to mention that in previous results is showed that

students do not trust in governmental organizations, as a result of corruption cases, in which it has been demonstrated that police organizations support organized crime. The IFAI[2] (recently named INAI[3]) is an agency moderately known by students. This is worrying because only the 39% of respondents affirm that they know the real functions of IFAI, the 37% knows about it but they don't know more about its functions and a 29% do not know anything about it. That could be due to low interest in young students to get information about their rights in privacy. . Outcomes are illustrated in the figure 4.

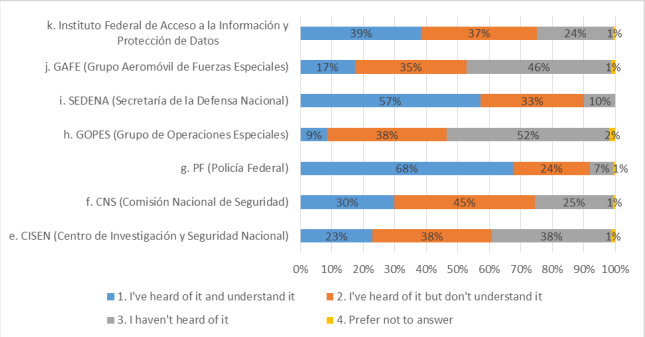


Figure 4. National Security Organizations

In the second section of the survey, the items were related to "threat to privacy" in ICT environments in both individual and organizations services. The first item in this section was "Do you feel that your use of the Internet involves taking risks with your privacy?". We found in this item that only 18% of students consider that the use of Internet involves a real risk in their privacy, while the 38% of respondents are convinced that the use of Internet do not suggest a reasonable risk in their privacy. 35% of respondents suggest that Internet affects in their privacy only to an extent. The remaining respondents 9% suggest that using Internet do not imply a risk in their privacy. The findings for this item show that students do not consider critical the topic of privacy when are browsing in Internet. That could be because young people are only focus in specific activities, for example emailing, social network and browsing for information [13], and it is probably that they don't mind about the risk over privacy that those activities imply. But the lack of knowledge in security makes cyber-criminals attacks a growing tendency. According with [19] there are some security topics in Latin-American and Caribbean zone: a) Increasing data breaches, b) Growth in targeted attacks, c) Social networking scams, d) Banker Trojans, and e) Segregation of malware and viruses.

A second item was developed in order to know how non-Internet activities involves risks in privacy. The outcomes for this item shows that only the 9% of the respondents indicate that activities who are not based in internet represent a real risk for their privacy. 17% of respondents have a moderate perception about the risk of their personal data in a non-internet activity. In contrast, most of people think that those kinds of activities do not have an important implication in their privacy, 48% selected "not so much" and the 25% answered "not at all". Only the 1% of the respondents prefers not to answer to this item. We point out that most people do not have interest about the risk of security in their non-Internet activities (more than 60% of respondents), in this respect college students don't take into consideration all the perspective in the security topic,. In Mexico the use of Internet for payments, shopping or trading is growing up in the last years [13] so it is easy to think that people do not take precautions when they are browsing or making transactions on Internet.

The report of the CIDAC[4] [2] shows eight of the principal crimes that Mexican population are afraid of, the Table 2 lists the crimes in order of importance.

Rank	Crime
1	Kidnaping
2	Intentional murder
3	Intentional injury with white weapon
4	Extortion
5	Robbery without violence to a passer
6	Robbery with violence to a passer
7	Robbery a vehicle with violence
8	Robbery a vehicle with violence

Table 2. List of main crimes in Mexico

We can observe in the list that at least two crimes can be related with the privacy information with or without the use of Internet: Kidnaping and Extortion. In Mexico most of the crimes associated with Kidnaping or Extortion uses the victims' personal information usually stolen from social networks or phone scams, etc., that's why the trust in the Government and Security Agencies has decreased year by year.

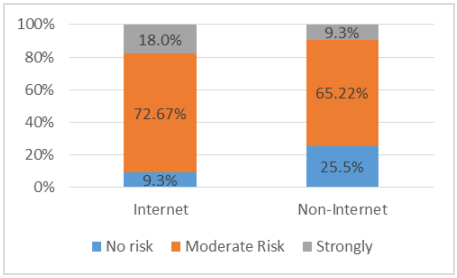


Figure 5. Risk recognition in the Internet and non-Internet activities

Regarding threaten of privacy, we collected some data about the perception of the Mexican students respect about some institutions. In this question we decide to divide the items in four sections: Individuals, profit organizations, non-profit organizations and government. The outcomes show in the first item that 57% of respondents consider that individuals close to them as possible threaten group to their privacy. Also Mexican respondents point that people that are close to them may not represent danger to their privacy or security with more than 60% as is shown in the figure 6.

[2] Acronim of Instituto Federal de Acceso a la Información y Protección de Datos.

[3] Acronim of Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales

[4] Centro de Investigación para el Desarrollo - (Research for Development Center)

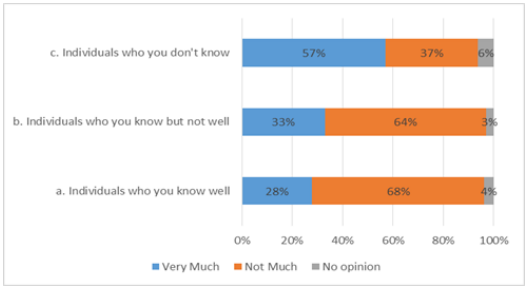


Figure 6. Threaten perception of individuals group

With the proliferation of technology many organizations (specifically telecom and computer) are fighting continuously to gain the marketplace. We have heard in the news the competition between Android and IOs, or Apple vs Samsung as very well-known examples. When someone buys a new mobile device, the first step is to configure the system creating an account in the developer system. In most of the cases users have to fill a form and send it to provider to enable the account to download apps, updates and more. The same topic is related for almost all operating systems in personal computers (Microsoft, Apple and some Linux distributions), where users regularly must to sing in with an account to have access in the system.

The findings in our survey suggest that Internet companies and telecom companies 76% and 63% respectively are perceived as attempting organizations in the privacy of users more than other profit organizations like computers companies (45%) or Software companies (43%) as is shown in the figure 8.

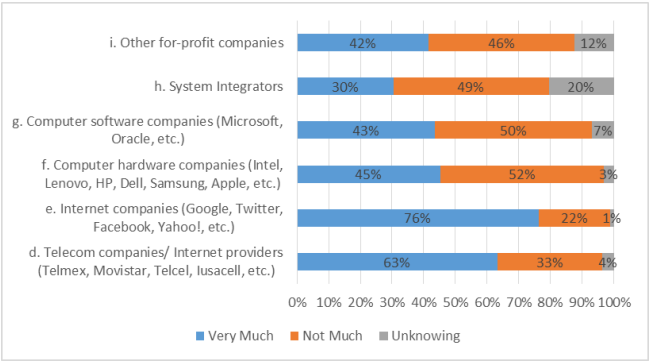


Figure 7. Threaten perception of profit organizations group

As well as there are profit organizations, there are non-profit organizations. In this aspect Mexican students were questioned about their perception of this kind of organizations. The survey indicates that 32% of people perceive that educational institutions have moderate influence in the violation to their privacy.

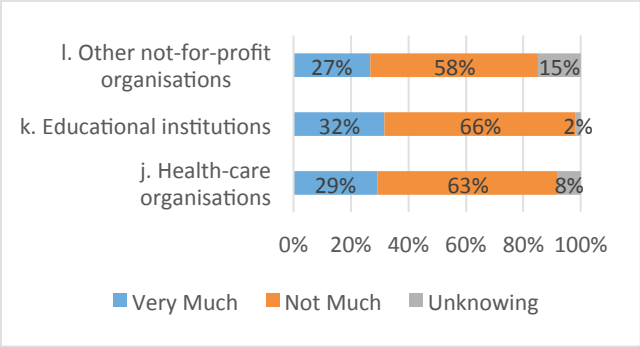


Figure 8. Threaten perception of profit organizations group

The final item shows the perception of the Mexican respondents about government institutions. The outcomes are important for this research because confirms the real situation in Mexico: the high level of corruption in the governmental institutions. The survey shows that 61% of respondents don't trust in Mexican Law enforcement and the 57% consider the secret services agencies threaten in their privacy or security.

Variable	Mean	Std. Dev.
Individuals who you know well	3.03	.974
Individuals who you know but not well	2.81	.769
Individuals who you don't know	2.33	1.043
Telecom companies/ Internet providers (Telmex, Movistar, Telcel, Iusacell, etc.)	2.12	.875
Internet companies (Google, Twitter, Facebook, Yahoo!, etc.)	1.86	.885
Computer hardware companies (Intel, Lenovo, HP, Dell, Samsung, Apple, etc.)	2.54	.966
Computer software companies (Microsoft, Oracle, etc.)	2.62	.974
System Integrators	2.68	.888
Other for-profit companies	2.60	.949
Health-care organizations	3.00	.865
Educational institutions	2.90	.950
Other not-for-profit organizations	2.84	.842
Law enforcement government agencies (Police)	2.16	1.061
Secret service government agencies (CISEN, CNS (e.g. PF, GOPES), SEDENA (e.g. GAFE))	2.11	1.090
Other government agencies (Health, Interior, Tax, etc.)	2.49	1.048

Table 3. Ranking of organizations that are viewed as a threaten to privacy

For the question number 4 in the survey, we decide to make a segmentation of each item for a better evaluation on the technologies that threat in the Mexican students' privacy. We divided the question in four segments: Home Technologies, Personal health, and online trading. Our findings show that GPS use is perceived as the most threaten home technology by 71% of Mexican students. This may be because nowadays GPS is a common technology used in cameras, mobile devices, watches, etc… more than other rising technologies such as smart meters or home automation. For personal computers and videogames, respondents pointed out that Smartphone in a 69% is the most perceived technology that attempts to their privacy over personal computers, with the 59% and videogame consoles (home consoles 29% and portable console 24%). According with [7] Smartphone had the 37% of penetration in Mexico in 2013, 76% of Mexicans don't leave home without the Smartphone and the key activities are browsing, listen music looking for address/maps and taking pictures. These findings are congruent because Mexican people use the Smartphone almost all the day and a one of the key

activity is looking for an address that imply the use of GPS, so that is why they feel vulnerable in their privacy. In relation with health technologies, it is important to underline that in Mexico not all the population have access to that kind of technologies. According to [8] technologies like T.V., mobile phones, Internet access, and informatics devices (PC, Software, etc.) have more penetration and usage than other technologies. In the case of Smartcards the survey points that 42% of Mexican students consider them as a threaten technology. The use of Smartcards for Mexican students is limited to transport card and don't have any other kind of smartcards. That could be the reason why 52% consider smartcards as a moderate threaten technology. However, people must take into consideration that the organization who develop the smartcard have access to user information like Personal Data, time of use average, Amount charged, etc.

Variable	Mean	Std. Dev.
Survey TV cameras	2.422	.995
GPS (Global Positioning System)	1.987	.960
Smart meter (an electricity meter providing your supplier with regular, approx. every 30 minutes, readings of your usage)	2.510	.815
Home automation which senses human activities (e.g., air conditioner, lighting apparatus)	3.013	.983
Personal computer (Widows machines, Mac, etc.)	2.291	.960
Smart phone (iPhone, Android, etc.)	2.000	1.000
Home video game console (Wii, PlayStation, XBOX, etc.)	2.878	1.042
Portable video game console (PSP, Wii-U, etc.)	2.904	.992
Smart card (transport card, gym card, etc.)	2.662	1.026
RFID (Radio Frequency Identification)	2.797	.882
Personal body monitoring (Fitbit, etc.)	2.604	.970
Behavioural targeting	2.922	.890
Social media services	2.375	1.002
Online shopping (Business to Consumer ecommerce)	2.236	.968
Online auction	2.464	.932
Online games	2.587	1.017
Making payments online	2.235	1.050

Table 4. Ranking of technologies that are viewed as a threaten to privacy

Although online trading activities are increasing in Mexico, not all citizens trust in them. According with [9] Mexico gain 9.2 millions of dollars in e-commerce activities growing 42% in 2013. The report also shows that the most used payment media was the credit card (used in 64% of online stores transactions). In the security context the report shows that 8 of each 10 online stores in Mexico provide to customer the store contact data and also offer secured access to sale webpage. In this respect the survey shows that these activities involve risk for privacy. 60% of respondents see payments as a privacy risk, and 61% for shopping. Social network is perceived as a less risky activity: 54% seen them as a privacy threaten.

4.1.2 Evaluation of Snowden's Activities in Mexico

The increasing use of Internet and communications services (phone calls, messaging, mailing, browsers, social networks, etc.) has opened several discussions about privacy, not only for organizations but also for individuals. Some years ago, Edward Snowden reveals some documents about massive spying and surveillance programs (PRISM and Xkeyscore).

In this section, we aim to show what Mexican students perceives about Snowden revelations. The survey shows that 50% of the students told that don't know about Snowden's revelations meanwhile 43% of the respondents sustain that they know something about him, and only the 7% of the young students responded don't know anything about the topic.

The survey also collects several opinions about Snowden's actions, so all the responses were grouped in order to understand the results as is shown in figure 9. In this analysis the 27.3% (44) of the respondents considered that the reason to leak information was to advise citizens about security and personal data protection. 23% (37) of the surveyed think that he leaked the information to inform citizen about the government procedures. As he worked as contractor for the NSA and CIA he had the opportunity to access specific documents about surveillance projects, in this way a minority respondents 5.6% (9) relates the obligation to inform with the moral and ethics consideration to inform about that kind of facts. However, a debate on ethics or morality is not enough, people should consider all the implications of this kind of facts, and how this implications may affect in globally context not only in security and protection of information.

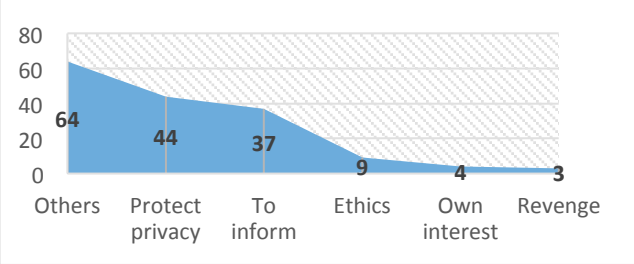


Figure 9. Why do you think Snowden determined to make those revelations?

4.2 Empirical Consideration about Influence of Snowden's Revelations

In this section we will evaluate empirically some specific data from the survey. In order to stablish a general supposition we consider to cross statistically some variables. To be congruent with the project objectives, we focused on behavior and the Snowden's Revelations survey.

This section will be focused mainly on four statements: 1) Does the concept of privacy and security have any influence on the users that are more aware of the risk involved in the use of the Internet?; 2) Users who rely on the good effects Snowden Revelations have had, are more compromised with doing the

same?; 3) Users more updated about WikiLeaks effects are more aware about the risk of privacy and security? and 4) Do the respondents who have selected social nets as the group which more threaten their privacy are more committed with changing the way of communicating?

4.2.1 Does the concept of right to privacy and security have any influence on the users that are more aware of the risk involved in the use of the Internet?

Methodology followed in the procedure for examining this research question was done making a division between the respondents they considered the use of Internet a risk or not (Q6) (Do you feel that your use of the internet involves taking risks with your privacy?) Then one specific question was empirically considered via T-test in which above these two groups were used as control and treat group. The question is whether the respondents who felt that the use of Internet involved risk with their privacy tend to validate privacy as important compared to those who did not consider the use of Internet as a threat. The result of the T-test (Table 5) reveals that the mean of the group who feels the use of Internet as a threaten for privacy (M=2.22, SE=.063) does not exceeds that of the group who doesn't feel (M=2.00, SE=.41) and the difference between these averages (DM=.225, 95% CI [-.534 ~ .985]) is a statistically significant at one percent significance level (t (144) = 0.59, p < .05). The results show that the null hypothesis is accepted, so it consider that those who attribute that the right to privacy is important to indicate that use of the Internet may harm their integrity.

Means of each group	Answer to Q6	yes	no
	Mean	2.22	2.00
	St. Error Mean	.063	.41
Difference between Means	Mean Difference	.225	
	95% CI	-.534 ~ .985	
Statistics	t-value	0.59	
	d.f.	144	
	p-value	< .05	

Table 5. Results of T-test: Q6 vs Q10

4.2.2 Users who rely on the good effects Snowden Revelations did are more compromise with doing the same?

The second statement is whether there is the tendency that respondents who rely on the good effects Snowden Revelations had in society would be more compromised to do the same in comparison with the group against this kind of behaviour. To confirm it average scores of Q28 (Have Snowden's revelations served the public interest or harmed it?) and Q30 (If you were an American citizen and were faced with a similar situation to Snowden, do you think you would do what he did?) are compared between the two groups. A first approach shows a difference in percentages of both groups. Figure 10 shows that the highest values in the affirmative answers correspond to those respondents who say they want to do the actions that Snowden did, so they consider it as good for the public interest, while those who responded negatively consider that harm the public interest.

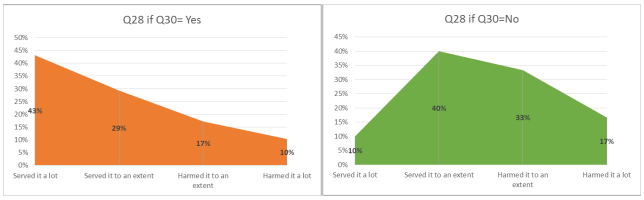


Figure 10. Differences between groups

According to the result of T-test, the respondents who think Snowdens´ revelations served for good purpose tend to feel a higher compromise with doing the same he did (M=1.95, SE=.133) compared to those who think what Snowden did harmed more than benefit in society (M=2.57, SE=.163), and the difference between the averages (D=.618, 95% CI [-1.055 ~ -.181]) is significant at one percent significance level (t (86) = 2.81, p < .05). The results shows that the null hypothesis is rejected, so in this item we have to considerate that the respondents who consider that Mr. Snowden's revelations is good for the public interest are more susceptible to do what Snowden did.

Means of each group	Answer to Q28	yes	no
	Mean	1.95	2.57
	St. Error Mean	.133	.163
Difference between Means	Mean Difference	.618	
	95% CI	-1.055 ~ -.181	
Statistics	t-value	2.81	
	d.f.	86	
	p-value	< .05	

Table 6. Results of T-test: Q28 vs Q30

4.2.3 Users more update about WikiLeaks effects are more aware about the risk of privacy and security?

The third statement was attempted to confirm whether the information spread in media about WikiLeaks (Q19) (Have you heard about Snowden's revelations?) served the population to be more aware about the implications privacy and security have got (Q6) (Do you feel that your use of the internet involves taking risks with your privacy?). The result of the T-test (Table 7) reveals that the mean of the group who had heard about Snowden revelations do not tend to be more aware about the risk of privacy and security (M=2.224, SE=.103) that of the group who did not heard about Snowden revelations (M=2.49, SE=.104) and the difference between these averages (DM=.244, 95% CI [-.534, .045]) is a statistically significant at one percent significance level (t (148) = 1.67, p < .05). The results shows that the null hypothesis is accepted, so it consider that those who had heard about Snowden's revelations not consider the privacy right as important as people that never heard about him.

Means of each group	Answer to Q19	yes	no
	Mean	2.24	2.49
	St. Error Mean	.103	.104
Difference between Means	Mean Difference	.244	
	95% CI	-.534 ~ .045	
Statistics	t-value	1.67	

	d.f.	148
	p-value	< .05

Table 7. Results of T-test: Q19 vs Q6

4.2.4 Do the respondents who have selected social nets as the groups which more threaten their privacy are more committed with changing the way of communicating?

The fourth statement was attempted to confirm whether the group of respondents that had chosen Social Nets as the more dangerous group that threaten their privacy (Q8-e) (How much do you feel Social Nets threaten your privacy?) had got a higher disposal for changing the way of communicating online (Q24) (Have you changed your way of communicating online since you heard about Snowden´s revelations?). The result of the T-test (Table 8) reveals that the mean of the group who had selected social nets as the groups which more threaten their privacy do not tend to make any changes in their way of communication (M=2.17, SE=.167) in comparison to the group who did not consider social nets as a threaten (M=1.84, SE=.098) and the difference between these averages (DM=.325, 95% CI [-0.157, 0.807]) is a statistically significant at one percent significance level (t (117) = 1.34, p < .05). We found this item as special indicator, because social networking is one of the most activities in Internet by Mexican citizens, In this respect, most respondents that considered change their way of communicating in Internet contemplate that social networking may attempt to their privacy, at the same time we can conclude that the other part of respondents may ignore that this activity may harm their integrity.

Means of each group	Answer to Q8-E	yes	no
	Mean	2.17	1.84
	St. Error Mean	.167	.098
Difference between Means	Mean Difference	.325	
	95% CI	-.157 ~ .807	
Statistics	t-value	1.34	
	d.f.	117	
	p-value	< .05	

Table 8. Results of T-test: Q8-E vs Q24

4.2.5 Do we approach Snowden´s Revelation depending on our gender?

In this item we aims to evaluate if there is a significant impact between variables Gender and Q28 (Have Snowden´s Revelations served the public interest or harmed it?). A cross table was elaborated in order to identify if there is a difference among gender and the impact of the Snowdens' revelations. The Table 9 shows that female consider that Snowdens' revelations are beneficial for the public interest a little more than the male respondents. We can conclude that there are no significant impact between both variables, that is both males and females have the same perception for Snowdens' facts.

Means of each group	Answer to Q1	Male	Female
	Mean	2.06	2.01
	St. Error Mean	.111	.129
Difference between Means	Mean Difference	.044	
	95% CI	-.292 ~ .380	

Table 9. Statistic Results: Q1 vs Q28

5. CONCLUSION

The controversial topic of surveillance, and threaten to privacy has grown in the last years as result of the statements of Edward Snowden. Mr. Snowden reported some actions of various government agencies in the United States of America, those acts included access to phone calls, access to emails and access to personal data of citizens of North America and the world. We can assume that these actions could be considered as the starting point for the world's population will be warned about the vulnerability of information. This vulnerability not only could be by Internet technologies, there are several methods and technologies to access it (i.e. ICs, radio frequency equipment, retina scanners, face recognition systems, etc.). In this respect, data protection can be individually, any user can prevent infiltrations in their private information through the use of protection tools (firewalls, antivirus, etc.) or a using a set of security policies, web browser privacy or social networking settings. However, when a service is obtained by an organization (Microsoft, Google, Facebook, etc.), privacy policies are applied on the side of the organization through a privacy contract, so this imply that most organizations could share, sell or do anything with the user data.

This study was intended to obtain an overall framework of the perceptions of young Mexicans in relation to Snowdens' case. In a first conclusion we could point out that Mexican students consider important their right to privacy and most of them have not a good understanding what privacy is. Another important finding for this research is that respondents affirm that they agree with Mr. Snowden actions and they could do that he did because this acts are beneficial for society and they consider to reply if necessary. However, outcomes show that Mexican students prefer not to bring personal data to some institutions like Internet companies, Telecom companies and government agencies, because in Mexico citizens are facing several social problems (corruption, illegal sales of private information, blackmailing, etc.). In relation, the findings shows that Mexican students are not informed about all the governmental security agencies in Mexico. The most known agencies were Policía Federal (PF) and Secretaría de Defensa Nacional (SEDENA). Today in Mexico a protection data laws[5] have been promulgated in order to regulate the security of the citizen's information, nonetheless the overall impressions suggest that most of the citizens have lack of knowledge of the laws and may have mistrust of them.

[5] The main law to data protection in Mexico is the Ley Federal de Protección de Datos Personales en Posesión de los Particulares, published in the Official Journal of the Federation on July 5, 2010

6. REFERENCES

[1] S. Aguayo Quezada, La Charola: Una historia de los servicios de inteligencia en México, Editorial Ink, 2014.

[2] V. Baz, R. CH, S. Aguilar, 8 Delitos Primero, Índice Delictivo CIDAC, México, D.F, 2013.

[3] V. Bello, La libertad de expresión y el PRI , PVEM y Panal, Zócalo Saltillo. (2015) 1–2.

[4] Cámara de Diputados., Constitución Política de los Estados Unidos Mexicanos, (1917).

[5] EFE México, México protesta tras las nuevas revelaciones de espionaje por parte de EEUU, El Mundo. (2013) 1–2.

[6] Excelcior, EFE, Cronología: Paso a paso del caso de los normalistas de Ayotzinapa, Excelcior.com.mx. (2015) 1–7.

[7] M.M.A. Google Inc., Ipsos, Our Mobile Planet, (2014).

[8] Instituto Nacional de Estadística y Geografía, Estadisticas sobre la disponibilidad y uso de tecnologia de informacion y comunicaciones en los hogares, 2013, 2013.

[9] R. Juárez, Estudio de Comercio Electrónico México 2013, México, D.F, 2013.

[10] A. Langner, Espionaje "revela que EU no confía en México," El Econ. (2013) 1.

[11] V. Lerner Sigal, Espionaje y revolución mexicana, Hist. Mex. XLIV (1995) 617–643.

[12] A. Marks, Drug Detection Dogs and the Growth of Olfactory Surveillance: Beyond the Rule of Law?, Surveill. Soc. 4 (2007) 257–271.

[13] P. Menéndez, E. Enríquez, Estudio sobre los hábitos de los usuarios de internet en México 2014, México, D.F, 2014.

[14] NTX / MIQC, Redes sociales , anzuelo para el robo de información, Informador. (2014) 2014–2016.

[15] L.G. Ornelas Núñez, M. Higuera Pérez, La autorregulación en materia de protección de datos personales: la vía hacia una protección global, Rev. Derecho, Comun. Y Nuevas Tecnol. (2013) 1–32.

[16] Proceso, Ordenan a la PGR abrir expedientes por genocidio en 1968 y 1971, Proceso.com. (2015) 1–5.

[17] Proceso, El Ifai inicia procedimiento de sanción contra Google México, Proceso.com. (2015) 1–5.

[18] J. Sánchez Onofre, El caso Snowden pasó inadvertido en México, El Econ. (2014) 1–5.

[19] Symantec, Tendencias de Seguridad Cibernética en América Latina y el Caribe, 2014.

Judging the complexity of privacy, openness and loyalty issues

Iordanis Kavathatzopoulos
Uppsala University
Box 337, SE-751 05
Uppsala, Sweden
iordanis@it.uu.se

Ryoko Asai
Uppsala University
Box 337, SE-751 05
Uppsala, Sweden
ryoko.asai@it.uu.se

ABSTRACT

Privacy protection and whistle-blowing are controversial issues. Privacy has to be protected but it hinders access to correct information. Whistle-blowing is necessary for correct decision-making, neutralizing wrong beliefs and preventing crime but it may destabilize groups, institutions and societies, and cause conflicts. The question investigated here was whether people judging the controversial issues of privacy and whistle-blowing take a moralistic or a philosophical approach. The hypothesis was that homogeneous responses point to a philosophical approach whereas responses correlated with cultural background point to a moralistic approach. Participants' responses to a questionnaire on Manning and Snowden cases did not produce an unambiguous picture, and this result did not lead to a decisive answer to our hypothesis question.

Categories and Subject Descriptors

K.4.1 [**Computer and Society**]: Public Policy Issues—*abuse and crime involving computers, privacy, use/abuse of power*

Keywords

Culture, decision-making, ethics, privacy, Edward Snowden, Sweden, Whistle-blowing

1. CONTROL OF INFORMATION, CORRECT THINKING AND DEMOCRACY

Privacy is fundamental to us since it seems to be a necessary condition for the independent existence of any organism or organization. Privacy belongs to a person even with or without being conscious about it. Privacy gives us integrity as person [1]. As Whitman summarizes, privacy is fundamental to our "personhood", especially in the western culture [2]. This is strongly depended on our ability to control information about ourselves perceived by significant others, e.g. rhetoric, cheating or mating behavior etc. Therefore we have to protect privacy.

Correct information, on the other hand, is also a fundamental condition for the independent existence of living organisms. It is also important for groups of organisms when they gather together in organizations, and when they need to make common decisions. Correct information is a necessary condition for correct decisions, lest the organism or the organization disintegrate. Therefore we have to go behind the façade of people and organizations to take a look on the unvarnished information hidden there.

We need to protect our privacy; and we need to breach the privacy of significant others in order to gain correct information about them, e.g. to foresee the behavior of others toward us, to know their true abilities and weaknesses, etc. This is an unavoidable contradiction regarding the principle of privacy which we have to consider, for example, in formulating policies, making decisions or taking actions. The balance between protecting privacy and access to information is significant not only for organizational decision making but for individual decision making as well. Privacy protection may prevent us from exerting our right to acquire knowledge in a fair and transparent way.

In making correct decisions we need correct information. Besides the need to bypass the efforts of others' to protect their information we also need to have an internal dialogue examining the correctness of the information we already have. This means openness to new ideas or tolerance (actually we need active promotion) of criticism of established truths. We need dissent voices. We need reports of things that are not correct. We need information about possible threats to important values in our groups and societies. Of course this questioning and disclosing process has to be inside the head of a right-thinking individual, but it is also necessary in a society. But by opening up for this dialogue in groups we run the risk of making it easier for significant others to access uncensored information about us. We make it easier for them to invade our privacy, according to the above line of argumentation.

Whistle-blowing is necessary for correct thinking and decision making. If anybody feels something is wrong one is obliged to report it to others so it can be judged in a dialogue. This is logically necessary, but in reality it causes problems, ironically for the same reason it makes it necessary. So here we have one more contradiction.

2. SWEDISH POLITICAL CULTURE AND CONTEXT

Sweden has diverse ethnicities inside the country. The survey participants have different cultural backgrounds. Although around 80% of the participants have Swedish nationality, some of them have grown up in families having different cultural backgrounds, and some have one Swedish and one non-Swedish parent. Suppose Swedish students having different cultural backgrounds and some 20% are foreign students, the survey results could have been reflected more or less by cultural differences compared with other surveyed countries. Anyway, the answers may be influenced by the Swedish political system and culture as it may have been for different cultural backgrounds.

It is important to consider some special characteristics of the Swedish society as a frame of understanding Swedish attitudes toward privacy and whistle blowing. Sweden is a homogeneous society, very conformist and well-disciplined valuing highly social cohesion and order. Collectivistic ideas are dominating, the independence of the individual is not so important. Family and friend connections in social, political and business activities play a minor role compared to the well-functioning of the overall societal norms, rules and beliefs. Political and moral correctness, of any time being, is directing what is going on and what people believe is right and wrong. Sweden is a typical Lutheran society where strengthening central authorities and maintaining trust on society as a whole is more important than questioning and doubting. As a society it is very sensitive regarding whistleblowing issues in other countries and cultures, and tolerant and supporting superficial or not so radical whistle-blowers domestically. However, in really daring cases where important political, financial or moral values are at risk Swedish society, state and public react intolerantly and in suppressive way (eg. Bofors affair, IB affair, Tsesis accident, radical political parties).

In Sweden, information and communication technology (ICT) has been highly permeated into society and people use Internet very actively not only at home/office but also through wireless mobile broadband, comparing to other OECD countries[3]. Freedom of Information Laws has been highly supported by the government and people can access broad public information officially. And public trust to the government is significantly higher compared to other countries, as we stated above. Basically people use credit cards and paying bills via Internet banking systems using home Internet, mobile phones and other applications. Swedish government aims to make society cashless, based on high public trust, low corruption and stable political procedures. Although people in Sweden are generally aware of the risk of breaching their privacy, usually they show relatively high trust on governmental institutions, especially organizations covering social welfare and education.

Sweden has been involved with Julian Assange case since 2010. He is being investigated about rape allegations in Sweden. Because of this situation, people might have been more interested in Wikileaks, Manning's case and Snowden's revelations than other surveyed countries.

3. HYPOTHESIS

The question explored in this paper is how people in Sweden react to the issues of privacy and whistle-blowing. University students from different backgrounds, cultures and countries have participated in a survey on Snowden's whistle-blowing case. How do students react to this case? How do they understand it? Do they grasp the complexity of the issue or do they judge the actions of Snowden and authorities according to their personal beliefs? Do they focus on the normative content of the story or on the nature of things and the way things happened? Do they see strengths and weaknesses with different alternatives? Do they consider in their judgments the inherent controversies of relevant conditions of privacy and whistle-blowing?

The way survey participants see the Snowden case could be discerned in a possible difference between cultures and backgrounds. A democratic, logical or philosophical approach should be independent of culture whereas a moralistic approach should be affected by the beliefs in different cultures.

4. METHOD

4.1 Questionnaire

The questionnaire contained totally 40 questions (29 multiple-choice questions, 10 open questions, and 1 combined (multiple-choice/open) question) about privacy awareness and whistle-blowing. Almost all questions had a free-answer column attached, and when the respondents chose the alternative "Other?Please specify" they could use the free-answer column. 6 of 29 multiple-choice questions and 2 of 10 open questions asked participants about their basic attributes (gender, nationalities, age, cultural background etc.). 10 questions were general questions about privacy (7 multiple-choice and 3 open questions). The rest were questions on a few cases (Wikileaks, Mannings and Snowden). The response rate of multiple-choice questions is around 72 percent and the rate of open questions is around 40 percent, except two questions asking respondents to write name and email address.

4.2 Participants

Undergraduate and graduate (Master and PhD) students who took courses related to IT and ethics in Uppsala University the last two years were invited by sending them an email with the link of questionnaire. The survey was conducted on voluntary basis. Respondents could participate in the online survey from 5th October 2014 through 11th November 2014.

The total number of respondents is 318 (male 228, female 87 and other 3). 44 percent of them belong to "25+ years old" category. Teenagers occupy around 7 percent of all participants. Most of the participants (92 percent) belong to Uppsala University, 6 percent came from other universities in Sweden and 2 percent from other universities in foreign countries. More than half of participants major in science and engineering while students in the liberal arts and humanities are few (8 percent). 253 of 318 participants were Swedish or mixed Swedish having Swedish nationality. The other cultural groups were as following: 8 Chinese, 6 German, 6 Greek, 5 Indian, 3 French and so on. Participants belonged to 30 different nationalities.

4.3 Procedure

The questionnaire was accessed online. The participants received information about the website of the questionnaire and they could respond to the questions any time they preferred during the period 5th October -11th November 2014.

The platform of the questionnaire was the free online survey software SurveyMonkey. All responses have been stored in a common data base from which they could be retrieved and analyzed.

5. RESULTS

5.1 General

In the survey, more than 70% of the participants talked about Snowden's revelations with others, and almost 50% of the participants changed the way of communicating online using systems. And also we can also observe their high trust in the society on the survey results about how to perceive Snowden's revelations. About 85% of the participants think Snowden's revelations served the public interest and the US government should not pursue a criminal case against Snowden. Moreover, around 50% of the participants say they would do the same as Snowden did if they were faced with a similar situation to Snowden in Sweden. Behind this high percentage, many participants believe the Swedish government would be more transparent and trustworthy than the US government, and information relating to the government and public interest should be open in public. Sweden's homogeneous political, ideological and value conditions seem to hinder participants awareness of the complexity of the whistle-blowing and privacy issues.

5.2 Gender

The results show that there is a strong correlation between gender and knowledge about the contents of Manning's and Snowden's revelations. Males reported a greater amount of knowledge about both Manning's and Snowden's revelations.

Table 1: Knowledge about Q18 Manning's (M) and SQ 24 Snowden's (S) revelations, percentage female and male

Gender	A lot M-S	A fair amount M-S	Not much M-S	Little M-S
Female	0%-0%	15%-22%	50%-63%	35%-13%
Male	5%-10%	47%-47%	31%-34%	17%-7%

Table 2: Whistle the blow like Q31 Snowden if American or Q34 Swedish citizen, percentage female and male

Gender	Yes American-Swedish
Female	25%-36%
Male	36%-51%

It is also clear that there is a correlation between gender and the willingness to follow Snowden's example as a whistle blower if the participant is an American or a Swedish citizen. Male participants were more willing than females to follow Snowden's example independently if they were thought to be Americans or Swedish citizens (**Table 3**).

Gender was significant regarding the importance of Manning's revelations. More males than females had a positive opinion about the impact of Manning's revelations on society. On the other hand there was no correlation to the opinion about the impact of Snowden's revelations.

Table 3: Correlations between gender, knowledge about Manning and Snowden, and willingness to follow Snowden's example as a whistle-blower

Gender	Q18 Knowledge Manning	Knowledge Snowden	American does as Snowden	Swedish does as Snowden
Pearson Corr.	**.317	**.284	*.148	*.142
Sig.(2-tailed)	.000	.000	.041	.050
N	182	192	191	190

** Correlation significant at .01 level
* Correlation significant at .05 level

Table 4: Gender and the importance of Q19 Manning's revelations, percentage female and male

Gender	Serve lot	Serve to an extent	Harmed to an extent	Harm a lot
Female	17%	46%	4%	0%
Male	43%	36%	4%	1%

Table 5: Correlation of gender and opinion about the impact of Manning's revelations.

Gender	Q19 Impact of lManning's revelations
Pearson Corr.	**.224
Sig.(2-tailed)	.002
N	182

** Correlation significant at .01 level

5.3 Age

The results (Table 6) showed a correlation between age and understanding of the privacy right. The older the participants are the more they understand what the right to privacy is. We have also to consider that the variation of age was only between 18 and 25+.

Table 6: Age

Age	Q14 Understand right to privacy
Pearson Corr.	**-.140
Sig.(2-tailed)	.038
N	219

** Correlation significant at .05 level

Age was not correlated to the opinion of how important the right to privacy is (Q11) or to any other of the questions.

5.4 Cultural backgraound

Participants from different self-defined cultural backgrounds ranked privacy and whistle-blowing issues rather homogeneously, even in questions where the answer alternatives were dichotomous, like in Q31. However it is not meaningful to test the significance of this summary picture of the results since the vast majority of the respondents defined themselves as North Europeans (see **Table 7**).

Table 7: Cultural background and teh rating of some important questions about privacy and whistle-blowing

Cultural Background	Q7 Risks Internet	Q11 Privacy meaning	14 Privacy right	Q18 Manning info	Q24 Snowden info	Q29 Snowden impact	Q31 American citizen	Q34 Swedish citizen	Q38 Privacy security
North European N=244	3(4)	4(4)	3(4)	3(4)	3(4)	5(5)	2(2)	1(2)	3(5)
South European N=11	3(4)	3(4)	3(4)	2(4)	2(4)	4(5)	2(2)	1(2)	3(5)
Latin American N=1	3(4)	3(4)	3(4)	1(4)	2(4)	1(5)		1(2)	4(5)
Arabic N=4	3(4)	4(4)	3(4)	4(4)	3(4)	4(5)		2(2)	3(5)
African N=1	3(4)	4(4)	2(4)	1(4)	2(4)	5(5)		2(2)	1(5)
South Asian N=5	3(4)	4(4)	4(4)	2(4)	2(4)	5(5)	2(2)	2(2)	1(5)
East Asian N=12	3(4)	4(4)	3(4)	2(4)	2(4)	4(5)	2(2)	1(2)	4(5)
Southeast Asian N=6	3(4)	4(4)	4(4)	2(4)	2(4)	4(5)	2(2)	2(2)	5(5)
Other N=23	4(4)	4(4)	3(4)	2(4)	3(4)	5(5)	1(2)	1(2)	3(5)

Ranking and in parenthesis number of alternatives.
If there are only two answer alternatives: 1=Yes and 2=No.

6. DISCUSSION AND CONCLUSIONS

The results are not unambiguous. It seems that our hypothesis that homogeneous answers of participants with different backgrounds has been supported, meaning that participants focus on democratic, logical or philosophical aspects when they judge the controversial issues of privacy and whistle blowing. Of course this may be true under the condition of non-democratic and non-philosophical approach being dependent on culture. If so, then participants from different backgrounds would give answers correlated to their corresponding cultures.

However if we define gender and age as relevant to culture we may see that there are some differences regarding knowledge about Manning's and Snowden's revelations, whistle the blow as Snowden did, the importance and the impact of Manning's revelations, and understanding of the importance of the right to privacy.

The main weakness of this study, regarding the role played by cultural background as a sign of the way (philosophical or moralistic) participants see the issues of privacy and whistle blowing, is the most of the participants define themselves as North Europeans. However, the present survey is part of a broader investigation taking place in many different countries. Since the same questionnaire has been used in different countries and cultures it is possible to compare the responses of the other surveys answering in a more confident way the above stated hypothesis.

7. ACKNOWLEDGMENTS

This study was supported by the MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan) Programme for Strategic Research Bases at Private Universities (2012-16) project "Organisational Information Ethics" S1291006 and the JSPS Grant-in-Aid for Scientific Research (B) 25285124 and (B) 24330127.

8. REFERENCES

[1] C. Fried. Privacy: A moral analysis. *Yale Law Journal*, 77:475–493, 1968.

[2] J. Q. Whitman. The two western cultures of privacy: dignity versus liberty. *Yale Law Journal*, pages 1151–1221, 2004.

[3] OECD. Oecd(indicator), doi: 10.1787/69c2b997-en, 2015. (Last accessed on 24th June 2015)

'That Blasted Facebook Page': Supporting trainee-teachers' professional learning through social media

Martyn Edwards
Sheffield Hallam University
m.edwards@shu.ac.uk

Dave Darwent
Sheffield Hallam University
d.darwent@shu.ac.uk

Charly Irons
Pontefract College
charlyirons@gmail.com

ABSTRACT
The creation and use of a Facebook group amongst trainee-teachers in post-16 and further education on a PGCE course at a large university in the North of England was studied. The Facebook group was self-initiated and self-managed by the trainee-teachers as a means of socialisation and peer-support amongst themselves. Data was gathered through parallel interviews with a PGCE trainee and a course tutor. Interviews were semi-structured using Tuckman's stages of group development (forming, storming, norming, performing) to explore the functioning of the Facebook group throughout the duration of the PGCE course. The role of teacher-trainers in influencing professional learning within the Facebook group initiated and owned by the trainee-teachers themselves was explored using the didactical triangle as a theoretical framework. It was found that the Facebook group was highly-valued both for supporting socialisation amongst trainee-teachers and as an additional means of mediating the course content of the PGCE. Lessons can be learnt both by trainee-teachers using social media for socialisation and peer-support and by course-tutors in designing teacher-training courses that may better ameliorate the pressures and sense of alienation trainee-teachers experience during initial teacher training.

Categories and Subject Descriptors
K.4 [**Computers and Society**]: Ethics

General Terms
Human Factors

Keywords
Teacher-training, social media, Facebook, socialisation, peer-support, didactical triangle.

1. INTRODUCTION
The issue explored in this article relates to the creation of a Facebook group by a cohort of 45 pre-service trainee-teachers enrolled on a PGCE course at Sheffield Hallam University, a large university in the north of England. Initially two distinct Facebook groups were created; one for group A and another for group B. Each Facebook group therefore comprised of approximately 23 trainee-teachers. They studied together in the University-setting for two days each week and for a further three days undertook a placement experience in a range of schools and colleges where they engaged in teaching practicum. The creation of the Facebook groups was initiated by the trainee-teachers themselves with its primary purpose being a self-managed initiative owned by the trainee-teachers to support socialisation and peer-learning amongst themselves. All trainee-teachers on the PGCE course were invited to join by those that created the Facebook group. It was a closed group for the trainee-teachers only and other stakeholders, including course-tutors and placement-mentors, were not invited to join the group.

After some weeks an incident occurred within the Facebook groups that created conflict. Whilst the full details of what occurred within the closed Facebook group are unknown (and it may be unethical to enquire as to what they were) some postings were made that were construed by others as 'unprofessional', or at least impolite. The postings were brought to the attention of a course-tutor who subsequently met with student-reps. It was agreed that the student-reps were best placed to address the issues that had arisen within the Facebook group and they were supported by the course-tutor to seek to establish a protocol for 'professional communication' to guide the future use of the Facebook group. Following this intervention no further incidents of conflict have come to the attention of course-tutors and the Facebook groups have continued as a self-managed initiative owned by the trainee-teachers to facilitate socialisation and peer-learning amongst themselves as intended.

There are considerable ethical problems that have prevented the researchers from seeking access to the posts made on the Facebook group as a source of data. These include issues of consent, confidentiality, power, and potential conflicts between the role of researcher and course-tutor. It has, however, been observed anecdotally that the Facebook group has been highly valued and seen as beneficial by a significant number of the trainee-teachers that have engaged with it. This inquiry seeks to understand what it is about the Facebook group that has been so engaging to some trainees, and whether it is the nature of the Facebook group as self-initiated and self-managed that has contributed to its popularity. It seeks to explore whether there is a role for teacher-trainers in supporting the development of professional attributes and professional behaviours in trainee-teachers through their use of social media, and if so what that role might be.

2. BACKGROUND AND CONTEXT
Professional standards play an important role in affirming teaching as a profession and qualified teachers as having demonstrated high standards of professional behaviour and

conduct throughout their training. The Teachers' Standards (Department for Education 2011) are used by head teachers in England to assess all newly qualified teachers in schools from their achievement of qualified teacher status (QTS) through to the completion of their statutory probationary period in schools. In the post-16 and further education sector newly qualified teachers are expected to demonstrate achievement of the professional standards set by the Education and Training Foundation (2014) for the award of Qualified Teacher Learning and Skills status, which has parity with QTS for schools.

An inspection of the two sets of professional standards (Department for Education 2011; Education and Training Foundation 2014) reveals that neither makes explicit reference to the ethics of social media use by teachers. Schoolteachers are expected to 'demonstrate consistently high standards of personal and professional conduct' (Department for Education 2011, p. 14). These 'high standards' are defined firstly by reference to authorities external to the teacher (statutory provisions; fundamental British values; ethos, policies and practices of the school in which they teach; and statutory frameworks which set out their professional duties and responsibilities). Secondly, they are defined in terms of what teachers must and must not do (observing proper boundaries; safeguarding pupils; not undermining fundamental British values; not exploiting pupils' vulnerabilities). The Education and Training Foundation, in setting out the standards for teachers and trainers in post-16 and further education, describe them as an aspirational document enabling teachers and trainers to take responsibility for their own professional learning (Russell 2014). These are articulated further as 'develop[ing] your own judgement of what works and does not work in your teaching and training'; 'develop[ing] deep and critically informed knowledge and understanding of theory and practice'; and 'develop[ing] expertise and skills to ensure the best outcome for learners' (Education and Training Foundation 2014, p. 2). Irrespective of the shift from prescription and proscription in the Teachers Standards to an increased emphasis on professional autonomy in the professional standards of the Education and Training Foundation, possibly reflecting the greater diversity of settings and specialist areas in the post-16 sector, the mention of social media is conspicuous by its absence in both sets of professional standards.

The use of social media such as Facebook has become ubiquitous amongst trainee-teachers and its use as an instructional medium has been the focus of a number of research articles (Ferdig 2007; Hramiak, Boulton & Irwin 2009; Wang et al. 2012; Goodyear, Casey & Kirk 2014; Soomro, Kale & Zai 2014). A distinction is often made in the literature between social network sites (SNSs) such as Facebook and learning management systems (LMSs) such as Blackboard and Moodle. Siemens & Weller (2011) identify the benefits of SNSs as encouraging peer-to-peer dialogue, promoting the sharing of resources, facilitating collaboration and developing communication skills. In contrast they describe LMSs as 'a fairly dry, bland set of communications that seems at odds with the forms of dialogue found in these spaces [SNSs] that mix humour, resource sharing, ideas, personal observations, professional updates and comments' (Siemens & Weller 2011, p. 166). In an analysis of over 68,000 Facebook posts made by university students, Selwyn (2007) found that SNSs were used mainly for social rather than academic purposes and that there was strong opposition to universities appropriating the use of SNSs for educational purposes which was seen as invading their social space and creating role-conflict for students who struggled with knowing what 'face' to project.

A particular dilemma for teacher-trainers is that of supporting trainee-teachers in their use of SNSs with the concomitant risks their use poses for teacher professionalism. This is one of the concerns that this article seeks to address.

3. DIDACTIC TRIANGLE AS A THEORETICAL FRAMEWORK

The didactic triangle (fig. 1) uses the student-teacher-content triad as a heuristic for analysing didactical situations (Bjuland 2012; Scoenfeld 2012; Jaako 2013) where the student, the teacher and the content are placed at the vertices of a triangle. Whilst it should be treated as a whole when used to analyse didactic situations it is common in practice to focus on pairs. The two-way arrow between the teacher and the student represents the pedagogical relation and is concerned with the practice of teaching, whereas the two-way arrow between student and content represents the didactic relation and is concerned with the practice of studying.

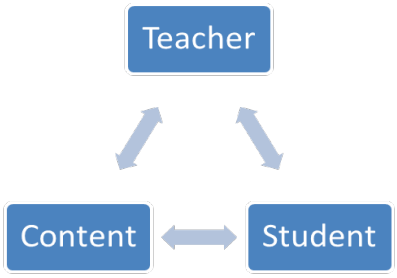


Figure 1: The didactic triangle.

Kansanen and Meri (1999) and Kansenen (2003) adapt the didactic triangle by introducing an arrow from the teacher to the line representing the didactic relation (figure 2). This fourth arrow is a one-way arrow and represents the teacher's efforts to influence the practice of studying so as to enhance learning. This is explained by Kansenen and Meri (1999, p.113):

> It is well known that teaching in itself does not necessarily imply learning. … If we describe the activities of the teacher as teaching, we would prefer to call the activities of the students as studying. It is this studying we can see and observe in the instructional process. … For the teacher, to bring about learning is the central task but to control the learning taking place is theoretically impossible. What the teacher is able to control, or rather to guide, is studying.

In relation to the present inquiry, the functioning of the Facebook group to facilitate socialisation and peer-support amongst the trainee-teachers could be seen as the didactic relation along the student-content arrow where the trainee-teachers act as the 'student' and the Facebook group is one method by which the 'content' is mediated. This inquiry is concerned with how teacher-trainers (in the role of 'teachers') might guide the didactic relation between trainee-teachers ('students') and the Facebook group ('content') so that professional learning is enhanced. In other words, it is concerned with how teacher-trainers might influence the process of 'studying' undertaken by trainee-teachers through the Facebook group.

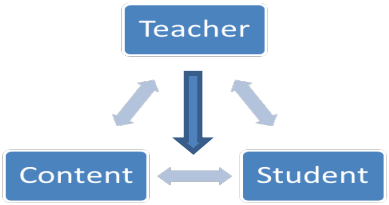


Figure 2: The didactic relation.

4. METHODOLOGY AND METHODS

Data was collected by conducting two parallel interviews. One of these interviews was conducted by a teacher-trainer with one of the PGCE trainees. This trainee was well-respected by her peers and represented a student-group at the staff-student course committee. She will be referred to throughout this article as 'PGCE trainee'. The other interview was carried out by another of the PGCE trainee-teachers who interviewed an experienced teacher-trainer responsible for course leadership of the PGCE. She will be referred to as 'course tutor'. Both interviews were audio-recorded and transcribed. The interviews were semi-structured using Tuckman's (1965) four stages of group development of forming, storming, norming and performing to explore the functioning of the Facebook group. Whilst the opinions expressed by the two interviewees cannot be taken as representative of those of the trainee-teachers or teacher-trainers more widely, they do provide an authentic voice of two influential individuals. Similarly, whilst this particular study into the use of Facebook on the PGCE course may not be transferable to other contexts, it is nevertheless of interest beyond its immediate context as an 'intrinsic case study' (Cresswell 2014) given the infiltration of social media into so many fields.

A key ethical consideration for ethnographic researchers in seeking to study culture-sharing groups is to negotiate access to the group prior to the research and to leave the field as undisturbed as possible after the research (Madison 2005; Ryen 2009). It has been discussed earlier that it would have been unethical to seek access to posts made on the Facebook page because of issues of consent, confidentiality and conflict of interest between the roles of researcher and teacher-trainer. It is also notable that recruiting participants for individual interviews also raised ethical concerns, particularly as the participants could be perceived, or perceive themselves, as speaking on behalf of others. There was a risk of disclosure during interviews of behaviour by individual trainees that could have been construed as 'unprofessional' that brought into question the fitness of those individuals to practice as teachers. This risk was managed through the process of participant recruitment by making it explicit and emphasising the importance of respecting confidentiality. The conducting of the interviews after completion of the taught content of the PGCE did alleviate some of the difficulties of leaving the field undisturbed since future attendance of the PGCE groups at University-based lectures were no longer being scheduled.

Each of the two interview transcripts were analysed independently by more than one researcher for key themes. These independent analyses were compared in order to enhance the reliability of the final analysis of the data. Once the analysis for each of the individual participants was completed the data was presented as a synthesis using Tuckman's (1965) forming, storming, norming, performing model to show the commonalities and contradictions between the viewpoints of the PGCE trainee and the course tutor on the PGCE course. The initial findings that emerged from the data were shared with the research participants at a group debrief session as a form of respondent-checking.

5. RESULTS

'Forming' is the first stage of Tuckman's (1965) model where the group acts as individuals and there is a lack of clarity about the group's purpose and the roles of individuals within it. This is followed by 'storming' where conflict arises as people begin to establish their place in the team. Storming then gives way to 'norming' where there is a level of consensus within the group, some clarity about individual roles and where group leadership emerges. Finally, 'performing' is where the group has a clear strategy and shared vision and is able to operate autonomously and resolve issues positively.

5.1 Forming

The creation of the Facebook group by the trainee-teachers to support socialisation and peer-support amongst themselves was discussed as early as the induction week of the course.

> **PGCE trainee:** There was like an induction week and then the week after there was an introduction to the course, and I think it had been mentioned in the induction week. But then it became something that was actually discussed properly in the introduction, perhaps in the break. And a couple of people said 'Should we set up a Facebook group so that we can kind of help each other out and ask each other questions and share our experiences from placement because we're only going to see each other sort of once a week?' And everyone thought that was a good idea and so one person decided to set it up and then invite everyone and that's how <pause> and then people invited each other if anyone was missed off because obviously you're not friends with everyone in the first place, you have to sort of befriend them first and then invite them into the group.

This description of the forming of the Facebook group as student-led, spontaneous, autonomous and organic contrasted sharply with practices from earlier cohorts of trainee teachers where socialisation and peer exchange was tutor-led, planned, managed and contained.

> **Course tutor:** We talk about ground rules [during induction week]. One of the ground rules and one of the expectations is about that we support each other, we're going to stay in touch with one another, we're going to have buddy groups, support groups. And whereas in previous years, 'Let's have a distribution list of names and phone numbers and email addresses', the past couple of years it's been 'Let's set up a Facebook group'.

An awareness of the potential for ethical issues to arise as a result of the creation of the Facebook group was anticipated from the outset by the course tutor, although this awareness was accompanied by a sense of powerlessness.

> **Course tutor:** Are we in a position to say 'No, you mustn't do it'? Cause they're going to do it anyway. [laughter]. So we know they're going to do it anyway, but let's be aware that they're doing it, and let's try and put things in place that can safeguard the integrity of the course, protect them and ensure that it's the supportive tool it should be.

It was evident that the reference to 'let's try and put things in place' did not extend to the imposition of rules, but rather to the establishment of a pastoral relationship in advance of anticipated issues likely to arise.

> **Course tutor:** I try very hard to establish a relationship with the group so that they feel that they have the opportunity to be open with me, to be honest with me, and to tell me things so that if things need to be dealt with, I'm able to do that. I also talk to the group about confidentiality. There's always a framework in which it has to work, but I try very hard to convey a respect for people who come and disclose and share and make sure that whatever they disclose or share with me is managed in a way that they feel happy with.

The trainee teachers, in contrast, did not anticipate the potential for ethical issues to arise from the use of Facebook in a professional learning context and ground rules established for the use of the Facebook group appeared to be arrived at tacitly, if at all.

> **PGCE trainee:** I don't know if we did any ground rules or not. Ground rules were mentioned at some point … and I think in the 'About' section of the group you can write something about what the group should be used for at least, but it wasn't like a list of rules or anything.

5.2 Storming

Several weeks after the creation of the Facebook group an incident that could be described as 'storming' took place. The interviewees recalled their experiences of the incident.

> **PGCE trainee:** I'm actually surprised that there wasn't more storming throughout the year. It only happened once, in the whole year, with all that kind of tension. And it's all been resolved now, there's no friction between the people that were involved. So I think it was just one person, then another person joined in saying that the things people were posting on the group were pointless, or something, and everyone else said, 'Well, you don't have to participate; it's voluntary and everyone else finds it voluntary and supportive, so why are you saying this?' … I have a feeling it might be because a couple of people were actually looking at the thread during [course tutor's] lesson, and she asked what was going on, and so we explained it to her <pause> and she was a bit concerned, and then at that point after the lesson one person sent the screenshot of some of the thread, and that was when [she] intervened and asked [the student reps] to come and meet her.
>
> **Course tutor:** The experience I've had is the experience I was expecting. That yes, … this virtual place that we're all very familiar with in our social lives, I'm very mindful that some of that, some of those social rules and that etiquette that's acceptable within a social Facebook world can sometimes leak into the Facebook page that's trying to operate we hope differently for this course. And so it was no surprise to me … that there would be conflict, that there would be upset, that there would be <laughter> issues. I didn't throw my hands up and go <gasp> 'What's going on?' I went, 'Oh, here we go'.

The ethos of the course tutor described earlier to 'try very hard to establish a relationship with the group so that they feel that they have the opportunity to be open with me, to be honest with me, and to tell me things' contrasted with the culture of the trainee-teacher group that regretted the involvement of the course tutor.

> **PGCE trainee:** There was no discussion of taking it to the [course tutor]. One person just thought, 'I'm going to take a screenshot of this and take it'. And everyone else thought that made us look very silly and childish, that we couldn't solve our own problems ourselves, and we needed the grown-ups to intervene. And that annoyed quite a few people because we thought, we could have resolved this. It was actually fizzling out at that point anyway.

5.3 Norming

The third stage of Tuckman's (1965) model is where consensus and agreement is reached on the purpose of groups and where agreements on individual roles and leadership emerge.

The view of the PGCE trainee expressed earlier that 'we could have resolved this' and 'actually it was fizzling out anyway' contrasted with the interventionist approach of the course tutor.

> **Course tutor:** I think that what happened was we then brought the course reps together with some of the team. We brought [another course tutor] in who could talk about the legality of it in terms of safeguarding and the dangers, but mainly about where we stood in terms of the law, and the University, and policy and practice. … and then we had the course reps who were excellent and what they did was they identified roles within themselves. … They identified roles and they identified actions that they took back to the group, fed back, and they operationalised it. They actually made it happen.

The course tutor stood back after the meeting with the course reps based on the assumption that they were going to establish rules with their respective groups and appoint a moderator to enforce those rules.

> **Course tutor:** I made a decision at that point to step back. They are beginning teachers. They have to acknowledge that they are now professionals. What and how they operated their social Facebook page is different to how they operate this Facebook page … there has to be different rules in place, and I, I'm assuming, that they did move into the next phase of working with the rules. If the rules weren't dealt with then I imagine the moderator would come in and said 'This is not acceptable.'

A very different perspective on the place of rules and the moderator were expressed by the PGCE trainee to those of the course tutor.

> **PGCE trainee:** I think the reaction against [the imposition of rules] would've been quite negative. Because we don't see it as something that's part of what we do in college time. It's part of our friendship group I suppose, and we don't, it's not something we want policing. We want it to be self-policing, and self-moderating. … [Ground rules] don't have to be really formal or anything, but just you know, what are our expectations of each other and how are we going to use the group, and let's make sure that we make that clear in the group description and that's enough. And then it's not imposed by the teachers, but the teachers are able to advise the students. I think that would be more helpful.

5.4 Performing

The final stage of Tuckman's (1965) model is where a clear strategy and shared vision emerges that allows groups to operate autonomously and resolve issues positively. It was apparent that this stage was reached where the Facebook group operated as an autonomous trainee-owned initiative.

> **PGCE trainee:** When we had a meeting as reps we discussed the fact that we should be a bit more polite to each other, and if we have any issues about the group, direct it to the group moderator and not just put it on the wall for everyone to see. But apart from that we've not had any incidents since so it must've worked, because that was over half-a-year ago <pause> and everyone has been treating everyone with respect since then.

A particular benefit of the Facebook group was the opportunity it afforded trainees to share some of the difficult experiences from placement, particularly given the lack of time in University to do so.

> **PGCE trainee:** … we've learnt a lot from each other I think. 'Oh, I've tried this in a lesson and it worked really well'; or 'I tried this and it was awful, it just fell completely flat'; or 'I used this piece of software, it was really good, it went down really well with my class'. That sort of possibilities to share ideas and share experiences, good or bad, we sort of wished that we'd maybe had a little bit of time to do a bit of group work on that rather than, not to replace what we were doing, but in addition to. … sometimes when we were asked to do something as a group we'd start talking about what we were supposed to be talking about [laughter] and then drift into what happened on placement that week and <pause> and we found that really helpful and supportive.

The immediacy of Facebook for sharing resources was acknowledged.

> **PGCE trainee:** I've just remembered something else we used it for as well is that when we did work in class, group work, we would photograph it. So we had one person who photographed it every week and posted it on the Facebook group … all you have to do is click upload from your mobile phone app, and it's just there instantly, so it was always there after the lesson.

A further benefit of the Facebook group was that it supported further socialisation amongst smaller groups of trainees away from the group site.

> **PGCE trainee:** Yeah, it brought us closer together you know. We maybe started communicating through our personal Facebook pages as well, as some of us have group private messages that we have going on where we maybe we don't want the whole cohort to hear our problems that we're having. So like three or four of us will talk about something that's going on at placement or a problem we're having with our coursework or something, offer support that way. But I think the group kind of led to that a bit more, because we sort of had to befriend each other on Facebook in order to invite each other to the group.

There was a vision for the continuation of the Facebook group after completion of the PGCE as a source of socialisation and peer-support into the trainee-teachers' NQT (newly qualified teacher) year.

> **PGCE trainee:** I think we all think it's going to carry on in some form and we hope that it'll be a way for us to keep in touch after the course is over. And still certainly in our NQT year I think we'll use it to share our experiences a bit and anything we've learnt from our NQT year. … I think we'll probably become less and less in need of it as we become more and more established in teaching and we rely more and more on our local networks in the colleges or schools we're teaching in <pause> but certainly you feel sort of, a bit of an outsider or loner at the start. Same when you're starting placement. Probably the same when you're starting your first job, and you, you're going through the same thing that everyone else is going through. So, you've got that shared experience of, you're all an NQT and you're sort of new to it <pause> finding your way.

6. DISCUSSION

The interview with the PGCE trainee revealed that the need for the Facebook group was identified by the trainee-teachers themselves from the very outset of their PGCE year. Regardless of the single storming incident referred to earlier, its effectiveness was confirmed throughout the year by the benefits derived by those who engaged with it. Furthermore, there appeared to be a widespread expectation amongst the trainee-teachers that the Facebook group would continue as a supportive network beyond the PGCE into the NQT year. A key benefit was to provide a sense of community and a way of sharing experiences that was not possible face-to-face because of teaching placement locations being dispersed across a wide geographical region. The need to ameliorate for the sense of alienation felt by the trainee-teachers on placement is not surprising where immersion in practice is the dominant model of teacher-training and trainee-teachers are located within a range of placement settings from the first week of their PGCE.

The operation of the Facebook group can be viewed as the process of studying along the didactic relation between student and content on the didactic triangle (Kansanen and Meri 1999). Interestingly, the operation of the Facebook group as a virtual community was not seen by the trainee-teachers as replacing the need for teacher-trainers to promote socialisation and peer-support through the University-based aspects of the PGCE, but rather emphasised that it was a missing aspect of their university experience that they would have liked more of. The importance of socialisation and peer-support, whether self-mediated or university-mediated, is attested to by Friesen and Besley (2013, p. 23) who argue that 'learning to *be* a teacher is as important as learning *how* to teach' [their italics]. Ticknor (2014, p. 291) argues that 'By reading, writing, talking, thinking, and interacting with others invested in the education community …, preservice teachers can engage in opportunities to negotiate professional identities within the supportive context of teacher education programs and build confidence as novice teachers'. Self-mediated support on social networks and structured opportunities for peer-sharing in the university aspects of teacher-training need not be mutually exclusive. Rather, the challenge is to more effectively embed socialisation and peer-support in teacher-training by strengthening connections between the different modes of support that exist.

Once the Facebook group had been established by the trainee-teachers it is not unsurprising that the teacher-trainers tried to influence its use. Such intentions are consistent with Kansenen

and Meri's (1999) assertion that whilst teachers cannot control the learning that takes place as a result of their teaching they can control, or rather guide, the process of studying. The course tutor's desire to guide the use of the Facebook group owned by the trainee-teachers was driven by the duty to safeguard the integrity of the course, protect the trainees, and ensure it was a supportive tool appropriate for professional learning. Whilst there was an implicit assumption from the course tutor that social media use was in her view inappropriate for teacher-training, the trainee-teachers appeared to hold a contrary belief that by locking down the security settings to create a 'closed group' any potential tensions between Facebook as a social space and a professional learning space were removed. Since teacher professionalism is a contested notion that is socially constructed it is surprising that the polarised positions of the course tutor and PGCE trainee were largely unchanged, and rather became entrenched, through their experiences of using the Facebook group on their PGCE year.

7. CONCLUSION

Several studies have been carried out into higher education students' personal use of social media and its impact on learning, and into universities attempts to appropriate these social spaces for pedagogic purposes. This present study focuses more specifically on the functioning of a closed group on Facebook set up by trainee-teachers in post-16 and further education as a self-initiated, self-managed and self-owned initiative to support socialisation and peer-learning amongst themselves at one university during their PGCE year. Given the ubiquitous nature of social media it seems highly likely that similar initiatives have taken place on teacher-training courses at other universities. This is an under-researched area and further studies are needed into the ways such groups are set up and the benefits and tensions that arise from their use in professional learning contexts, both for trainee-teachers themselves and the teacher-trainers responsible for their professional formation.

The Facebook group met a real need brought about by the trainee-teachers' sense of alienation from each other within an immersion-in-practice model of teacher-training and appeared to succeed in meeting that need for those that engaged with it. Nevertheless, the trainee-teachers still craved opportunities to share their placement experiences within the university-based aspects of their teacher-training. The challenge for teacher trainers is neither to discourage the use of social media nor to seek to control its use, but rather to create links between the informal social learning taking place on social media and elsewhere and the more structured parts of teacher-training.

It is suggested that increased opportunities for socialisation and peer-learning within teacher-training courses may mitigate against the risks of some trainee-teachers being disengaged from social media in a similar way to that in which social media use has compensated for the sense of isolation within immersion-in-practice models of teacher training for some trainees.

Professional standards provide little explicit guidance to trainee-teachers on social media use and practices differ widely across different educational providers. Wider ethical issues within the professional standards can be related to trainee-teachers' social media use including promoting diversity and inclusion, building collaborative relationships with colleagues, and operating within an ethic of respect. Assumptions about the ethical use of social media held by trainee-teachers themselves and teacher-trainers are likely to be challenged by the increased use of social media in professional contexts for different purposes.

8. REFERENCES

[1] Bjuland, R. (2012) The mediating role of a teacher's use of semiotic resources in pupils' early algebraic reasoning. *ZDM Mathematics Education*. 44, 665-675.

[2] Cresswell, J. (2014) *Educational research: planning, conducting and evaluating quantitative and qualitative research*. Harlow, Pearson Education Limited.

[3] Department for Education (2011) *Teachers' standards: guidance for school leaders, school staff and governing bodies*. [Online]. Accessed at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/301107/Teachers__Standards.pdf.

[4] Education and Training Foundation (2014) *Professional standards for teachers and trainers in education and training - England*. [Online]. Accessed at http://www.et-foundation.co.uk/wp-content/uploads/2014/05/4991-Prof-standards-A4_4-2.pdf.

[5] Ferdig, R. (2007) Editorial: examining social software in teacher education. *Journal of Technology and Teacher Education*. 15 (1), 5-10.

[6] Friesen, M. & Besley, S. (2013) Teacher identity development in the first year of teacher education: a developmental and social psychological perspective. *Teaching and Teacher Education*. 36, 23-32.

[7] Goodyear, V., Casey, A. & Kirk, D. (2014) Tweet me, message me, like me: using social media to facilitate pedagogical change within an emerging community of practice. *Sport, Education and Society*. 19 (7), 927-943.

[8] Hramiak, A., Boulton, H. & Irwin, B. (2009) Trainee teachers' use of blogs as private reflections for professional development. *Learning, Media and Technology*. 34 (3), 259-269.

[9] Jaako, J. (2012) Controlling the didactic relation: a case in process engineering education. *European Journal of Engineering Education*. 39 (4), 448-462.

[10] Madison, D. (2005) Critical ethnography: method, ethics and performance. Thousand Oaks CA, Sage.

[11] Russell, D. (2014) *New professional standards launched today*. [Online]. Accessed at http://www.et-foundation.co.uk/ceo-blog/new-professional-standards-launched-today/.

[12] Ryen, A. (2009) Ethnography: constitutive practice and research ethics. In D. M. Mertens & P. E. Ginsberg (Eds.). *The handbook of social research ethics*. Los Angeles CA, Sage.

[13] Schoenfeld, A. (2012) Problematizing the didactic triangle. *ZDM Mathematics Education*.

[14] Selwyn, N. (2009) Faceworking: exploring students' education-related use of Facebook. *Learning, Media and Technology*. 34 (2), 157-174.

[15] Siemens, G. & Weller, M (2011) The impact of social networks on teaching and learning. [Online monograph]. *Revista de Universidad y Sociedad del Conocimiento (RUSC)*. 8 (1), 164-170.

[16] Soomro, K., Kale, U. & Zai, S. (2014) Pre-service teachers' and teacher-educators' experiences and attitudes toward using social networking sites for collaborative learning. *Educational Media International*. 51 (4), 278-294.

[17] Ticknor, A. (2014) Negotiating professional identities in teacher identity: a closer look at the language of one preservice teacher. *The New Educator*. 10, 289-305.

[18] Tuckman, B. (1965) Developmental sequence in small groups. *Psychological Bulletin*. 63, 384-399.

[19] Wang, Q. et al. (2012) Using the Facebook group as a learning management system: an exploratory study. *British Journal of Educational Technology*. 43 (3), 428-438.

Carey Grammar School – a case study of the degree to which a digitally rich school can be considered to have the attributes of a digital society

Richard Taylor
International Baccalaureate
Malthouse Avenue
Cardiff
UK
44 29 20 54 77 47
richard.taylor@ibo.org

Michael Fitzpatrick
Carey Baptist Grammar School
4 Norma Court
Viewbank 3084
Australia
61 422 586 527
michaelgfitzpatrick@gmail.com

ABSTRACT
Many senior managers in schools see the acquisition and implementation of new information and communication technologies (ICTs) as key drivers in the advancement of their schools. Over the last 20 years pioneering schools, such as Carey Grammar School (Carey) in Melbourne, Australia have invested heavily in ICTs and created communities of digitally rich individuals. However, the extent to which these schools can be considered to have the attributes of a digital society, where there are shared beliefs, policies and practices with respect to the use of ICTs, has not been fully explored.

As schools such as Carey continue to heavily invest in ICTs, perhaps the biggest challenges senior managers will need to resolve are the apparent tensions that exist between the values and attitudes of staff, students and parents populations towards the ever changing digital environment. While it may never be possible for these schools to completely exhibit the characteristics of a digital society, it may be possible to measure the extent to which these schools shares common values and attitudes, and use the findings to inform future digital strategies.

To determine the extent that Carey is progressing towards becoming a digital society a questionnaire will be circulated to students in the 11th Grade (16 – 17 years old) that seeks to present their values and attitudes which underpin their relationship with technology. The analysis of the quantitative information from these questionnaires will provide baseline information that can be used to develop digital strategies in the future.

Categories and Subject Descriptors
K.4 [**Computers and Society**]: Ethics

General Terms
Management, Documentation, Human Factors

Keywords
Acceptable IT use policies, code of ethics, constrained ethics, digital hierarchy, digital immigrants, digital natives, digital society, digital wisdom, information society, negotiated ethics, networked environment.

1. INTRODUCTION
This paper is based on the initial investigations about the extent to which Carey may be considered as a digital society. This investigation will be carried out by testing a hypothesis linked to each of the three requirements of a digital society, the cultures of the tools, values and norms. The analyses will lead to tentative conclusions about the extent the school is a digital society and could provide a suitable baseline for any further studies.

2. A CHANGING TECHNOLOGICAL ENVIRONMENT
The technological environment of the second decade of the 21st Century is a significantly different one to that in the previous decade. The networked environment that students are familiar with is a far cry from that where computers were largely confined to a small number of rooms with a few stand-alone machines scattered throughout the school. Ubiquitous computing (usually using a variety of devices) is now the norm and the distinctions between home and school and personal or private use have become increasingly blurred. In this environment policies and practices that were developed more than a few years previously are likely to be anachronistic and/or ineffectual.

In this ever changing digital environment citizens will need to acquire more than just a range of key skills and competencies that may have been appropriate nearly twenty years ago when the first attempts were made to identify the characteristics of this rapidly evolving information technology based society. The IBM Community Development Foundation Report, "The Net Result -

Report of the National Working Party for Social Inclusion" [5] defined an information society as possessing the following characteristics:

- A high level of information intensity in the everyday life of most citizens, in most organisations and workplaces.
- The use of common or compatible technology for a wide range of personal, social, educational and business activities.
- The ability to transmit, receive and exchange digital data rapidly between places irrespective of distance.

Although the term information society is a term generally applied to society as a whole, the same principles can be applied to smaller 'societies' such as Carey. However, being considered an information society does not take into account the degree to which the behavioural values and attitudes of a school community as well as the procedures and policies are moving towards agreed standards.

It is proposed that an information society may be considered as being a precursor to a digital society. Based on research in South Korea, widely considered to be the most "connected" country in the world, Kwon et al suggested that a digital society is characterised by three cultures:

- Digital tools which allow humans to maintain his/her social life in the digital society and considered as the ground for other elements of the digital culture.
- Digital values which form a belief system that provides meanings or goals for human behaviours or social activities in the digital society.
- Digital norms which represents normative procedures and rules that are socially acknowledged in carrying out digital activities [6].

3. THE CURRENT SITUATION AT CAREY

The Carey school communities consist of approximately 1450 students (aged between 11-18 years old) as well as approximately 200 teaching and ancillary staff. Carey is fortunate as they have a parent community able to pay for the rent of hardware such as laptop computers. This additional support has allowed a more rapid progression towards the school becoming a "one-one"[1] laptop school.

To ensure the appropriate use of computers, Carey has developed a range of acceptable IT use policies. This is supported by a number of meetings such as when students enter from the feeder school where it is outlined how Carey's acceptable IT user policies work. This includes mentoring schemes where staff address issues that may occur from inappropriate digital behaviour and the development of a school culture of self-reflection where it is hoped students and staff will regularly evaluate their own digital behaviour. Throughout the school there is a positive reinforcement of digital behaviours to enable students to become responsible digital citizens.

[1] "one-one" laptop school means that each member student and member of staff has access to a laptop for use in the school

The definition of a digital society proposed by Kwon et al may be adapted for the study of Carey to provide the following three hypotheses:

Hypothesis 1: Digital tools

The introduction of new technologies (digital tools) at Carey as a positive agent of change is through a careful and considered, perhaps a slow tech, [9] approach

Hypothesis 2: Digital values

Ethical decision making, predicated on positive reinforcement, relating to the use of technology at Carey is seen as a result of a common belief system (digital values) such as the attitudes of different stakeholders in the school community about acceptable levels of privacy, anonymity, security and participation

Hypothesis 3: Digital norms

The policies and practice (digital norms) at Carey reflect the common procedures and rules for carrying out digital activities.

Using these cultures as success criteria digitally rich schools, such as Carey, can be seen to possess the culture of the digital tools; an environment where every student and teacher has constant access to a digital device and uses digital tools. However, the extent to which it applies to the other two cultures; the digital values and digital norms, is much less obvious and has not been fully explored.

4. DETERMINING THE EXTENT CAREY IS A DIGITAL SOCIETY

For Carey to be considered as a digital society it will be necessary to determine the extent to which members of the school community are digitally empowered and able to make free choices about their digital behaviour. Furthermore it is based on an assumption that the physical world and digital world of members of the Carey community have become inseparable [8].

The digital progression hierarchy, see Figure 1 below, seeks to create a framework that attempts to represent the progression from a technology rich community to a community of shared values and norms in where ethical decision making, "the why, how, when, where, and who of our digital footprint in today's world" [8] , lie at the core of this networked environment.

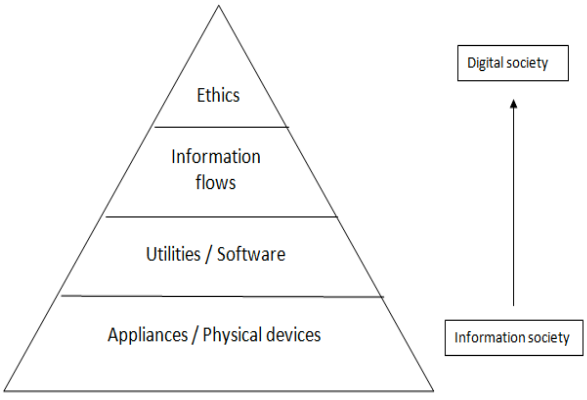


Figure 1: The digital progression hierarchy

In addition to the digital progression hierarchy the culturally negotiated ethical triangle [9], see Figure 2 below, can be adapted

to incorporate an information society and digital society. The assumptions made are that an information society will be characterised by restrictive policies whereas a digital society will be based on agreed behaviours or codes of ethics. Consequently, as the school progresses towards a digital society it is likely that more and more decisions pertaining to the use of ICTs will be based on negotiation and the position of the school will be seen to 'ascend' the triangle.

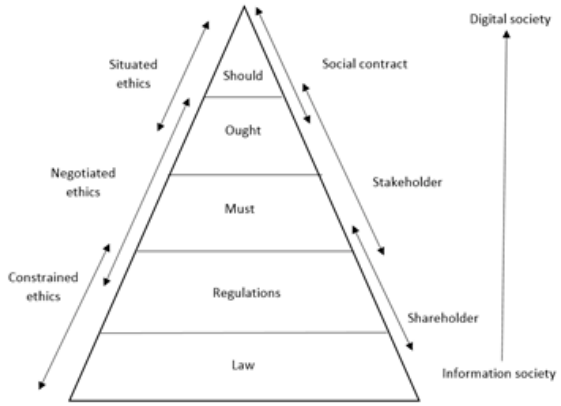


Figure 2: Locating information societies and digital societies in an updated version of the culturally negotiated ethical triangle

By analysing the information obtained from the questionnaire and making reference to the two models, it will be possible to position Carey on the continuum between an information rich community, which it has reached with its "one-one" laptop school policy, and a digital society.

Methodology

To gather information about this progression of Carey towards a digital society a questionnaire were circulated to 154 students in the 11th Grade (16 – 17 years old) and 61 staff that addressed a range of values and attitudes towards their relationship with technology. Carey was chosen for this study because a statistically representative sample of the values and norms can be obtained without excessive cost (time and/or effort), the senior management team are receptive to new ideas that may emerge from the research and the staff and students were willing to be involved in the survey. Despite the relatively short time the questionnaire was available online, three weeks, 67 students (44%), 32 staff (51%) and 30 parents (approx. 15%) responded.

The questionnaire had two foci. One was the relationship between the member of the Carey community and the digital technologies they use, this information made it possible to tentatively position the community on the models described earlier. The second was an evaluation of the effectiveness of the Carey Link facility[2] as this is the preferred mechanism for the flow of information within the Carey community.

As this research is focused largely on school age students, the anonymity of the students was the main priority. This was done by ensuring the information was collected transparently[3], the data collected was proportionate with the requirements of the research [4] and access to the data collected was confined to as few people as possible.

Throughout the development of the questionnaire the senior managers at Carey were kept informed so that they could help shape the nature of the questions. This would enable them to use the findings as part of their future planning.

5. FINDINGS

Figure 3 below shows the headline results from the questionnaire. Although three groups (students, staff and parents) were asked, only the results from the students and staff have been analysed. The sample of the parents will be analysed in the future.

Figure 3: Percentages of staff and students who answered "yes"

	Question	Student	Staff
1	Do you use the same device(s) at home and at School?	92	53
2	Do you use cloud platforms such as Google Drive or iCloud?	57	53
3	Do you regularly use social networking sites such as Facebook and Twitter?	90	53
4	Do you regularly use email, messaging services or other ICT services to communicate with family and friends?	82	87
5	Do you regularly evaluate and change your online behaviour on social networking sites?	40	35
6	Do you believe your online profile is an accurate reflection of yourself?	85	55
7	Do you regularly contact others in the Carey community using ICT?	91	100
8	Do you believe that it is important to separate your school role from your home role with regard to your ICT use?	57	84
9	Do you use Carey Link on a regular basis?	67	94
10	Do you think the design of Carey Link suits its purpose as a community communication tool?	74	55
11	Do you think that the use and development of ICT has improved the quality of communications between members of the Carey community?	85	83

The results largely mirrored anecdotal discussions prior to the questionnaire being circulated; students tend to have less physical

[2] Carey Link is the school's secure method of communicating between members of the school community. Users require a username and password to access Carey Link.

[3] A survey was set up using Survey Monkey that was only made available on the School Intranet. No personal data was collected.

separation from their digital devices, use social media more frequently, are less reflective about their digital persona, use Carey Link less and are less critical of its effectiveness.

The outcomes of the questionnaire were also linked to the digital progression hierarchy and the culturally negotiated ethical triangle. This was done in an attempt to provide a framework for the positioning of Carey between an information society and a digital society. From this baseline position subsequent studies could be carried out to determine the progression towards a digital society. Therefore the further Carey "moved up" the models, the more it can be considered as a digital society.

The results suggest that the community is a technology rich community using a wide range of application software. The high percentage of students (92%) who use the same device at home and school may be largely for economic reasons, but for many staff this may be for convenience such as file management. The use of the same device would suggest that for many home and school/work are a merging seamlessly into a networked environment.

In contrast there are differences in how students (85%) and staff (55%) see their online persona as an acute representation of themselves. This may be for a number of reasons and may be age related. It is also evident that both students (40%) and staff (35%) are less likely to be reflective about their digital persona. This may indicate that there is further development of the Carey community in its reflexivity and digital behaviour, but how this is done may require a more open approach to the technology.

The information about Carey Link suggests that the system is functional and does what is required. The students are less critical of it, but this may reflect the fact they place less demands on the system than staff. However, more research is required into the nature of Carey Link, and what is it trying to achieve before meaningful judgements can be made about its effectiveness.

6. ANALYSES

The analyses will relate the findings form the questionnaires to the three hypotheses.

Hypothesis 1: Digital tools

Both the staff and students believe that the use and development of ICT has improved the quality of communications between members of the Carey community. The careful and considered well-resourced introduction of new technologies (digital tools) through the "one-one" laptop scheme has been a positive agent of change. Carey is an example of a successful implementation of ICTs into a school, over a period of 18 years, being able to remain both at the cutting edge and using a slow tech approach.

Hypothesis 2: Digital values

Ethical decision making, predicated on positive reinforcement, relating to the use of technology at Carey is seen as a result of a common belief system (digital values) such as the attitudes of different stakeholders in the school community about acceptable levels of privacy, anonymity, security and participation

In the future it is likely that decisions at Carey will become increasingly focused on policies and practices and linked to establishing the extent to which shared values exist between different members of the Carey community. While it is accepted that some decisions will be as a consequence of constrained ethics (must), for example to filter inappropriate content to students on the school network, there could be a transition towards negotiated ethics (should, ought) as the basis for determining the appropriateness, or not, of applications such as Google Docs, Facebook and Twitter. The findings form the questionnaire suggest that there are some shared values, such as the use of collaborative tools such as Google Docs, but the thorny issue of social media still remains. It is also apparent that only a relatively small percentage of staff and students are continuously evaluating their online profile.

One mechanism that could be used to achieve shared values may be to implement a code of ethics at Carey. While it may not be possible to fully achieve this, the dialogue in involved in the process of trying to achieve this goal may prove be more productive than its implementation [1]. Furthermore the challenge of frequent technology changes may only enable any such code of ethics to be based on high level principles rather than specifically tied to practices.

Hypothesis 3: Digital norms

The policies and practice (digital norms) at Carey reflect the common procedures and rules for carrying out digital activities.

If there is a belief that many of the policies tend to be punitive [3], this may be indicative of a lack of trust towards technology and in particular the Internet safety, the overriding goal may be on harm reduction rather than attempting to positively influence values and norms [1]. This may require a shifting in two key areas; the perception of the Internet as a force for good and the need for a greater understanding of concepts such as privacy and anonymity. One aspect in the development norms will be the need to acquire the wisdom the make informed decisions about how ICTs can be effectively used to harness their potential. In the networked world we live, "our understanding of wisdom need to include an understanding of the digital world" [13].

7. CONCLUSIONS

The initial findings from the questionnaires completed by the Carey community suggest that it has moved beyond an information society and is progressing towards becoming a digital society. There is a well-developed technical infrastructure with all members of the Carey community having access to a laptop. There is the opportunity to use a wide range of application software, although in the case of social media this is more constrained. Within and beyond the Carey community there is a relatively free flow of information, with the possible exception of social media. The Senior Management Team are also taking a proactive approach to the digital education and empowerment of the Carey community. From the analyses it is possible to tentatively place Carey approximately half way "up" both models and this may be considered as the 2015 baseline position. However it was apparent that in any subsequent analysis of Carey the use of quantitative data would need to be superseded by qualitative data. This more in-depth analyses of respondents would be necessary to develop an accurate picture of their digital values and norms, but gathering this information would be considerably more time consuming and difficult to draw conclusions from.

The digital revolution has led to a change in what we need to achieve in terms of: confidence; self-belief; empowerment; resilience etc. Furthermore, the increasing morphing of digital and physical worlds may render attempts to distinguish digital policies from other policies at Carey redundant. Where it was

previously possible to write policies that made this distinction, will the networked world that the Carey community inhabits lead senior managers to ask, does our understanding of wisdom need to include an understanding of the digital world? And if so, how can policies and practices be developed with this in mind? How can we continue move the teaching of Internet safety away from a smoking-related safety approach (i.e. Can't be our fault, we don't allow it here – must have learned it at home) to a swimming-based approach (i.e. the more experience and knowledge you have, the safer you will be)? We need to move from an emphasis on digital literacy towards digital empowerment, where people of all ages, individually and collectively, are able to harness digital tools to enhance their lives and the lives of others [13].

It is unlikely a school based online environment, such as at Carey, can ever be completely unregulated. However, for Carey to exhibit the characteristics of a digital society there should be a greater propensity in decision making to be based on negotiated ethics. This could be considered as part of the further development of the successful mentoring and counselling strategies already in place.

The interchange of information on a free and casual basis at Carey is variable. This can be linked to the values and attitudes towards web based platforms such as Google Docs which are encouraged and social media, which tend to be discouraged. This is an area where future discussions about attempting to reach shared values and behavioural norms could be most productive

Within the Carey community, there are different perceived benefits and threats of new technologies between students and administrators, but they are not significant. However, it is noted anecdotally that staff perceptions of the benefits and threats of new technologies are not age related.

Looking ahead, it can be inferred that Carey's digital usage policies are moving towards a code of ethics approach, but is this providing a framework that is robust enough to evolve as the ICTs evolve? The concept of a digital society may be dynamic, so what constitutes a digital society may also change over time as the cultures of the values and the norms may coalesce. Carey will need to keep abreast of these developments in the digital landscape and attempt to develop strategies that can adapt accordingly. Will the senior managers at Carey be able formulate effective policies in such a rapidly evolving environment?

It is likely that the differing perceptions of privacy, anonymity and security may lie at the heart of any future discussions about the relationship between members of the Carey community and the digital technologies that are used. It may also be necessary to include empowerment to this list of terms. Currently the lack of empirical information linked to the distinctions (or not) between what these terms mean may not provide senior managers with either sufficient meaningful information to justify future digital strategies or to use generic terms such as "use the computer wisely".

Carey is clearly a pioneer in the adaptation of ICTs. The senior management of the school have kept it at the forefront of technological innovation. The next challenge is to strike an appropriate balance between empowerment, participation, security and regulation. It is not about the technology, nor the technological imperative [2], it is about behaviours and norms [7]. To do this effective education lies at the core.

8. ACKNOWLEDGMENTS

Thanks to the Carey community in allowing the research to be carried out in the school.

9. REFERENCES

[1] Burmeister O., 2013, *Ethical space*. The international journal of communication ethics Vol 10, No 2/3 2013

[2] Chandler D., 2008, Technological or media determinism. DOI= http://www.aber.ac.uk/media/Documents/tecdet/tdet07.html (last accessed 6 July 2015)

[3] Gotterbarn, 1996. *Software engineering: the new professionalism.* The Professional Software Engineer, C Myer ed., Springer-Verlag, New York

[4] Gotterbarn D., 2013, *Proceedings of the British Computer Society Forum, London*

[5] IBM Community Development Foundation. 1997, *The Net Result*. *Report of the National Working Party for Social Inclusion*.

[6] Kwon et al, 2013. *Index development and application for measuring the level of digital culture. Proceedings of the IADIS multi-conference on ICT Society and Human Beings, ed P Kommers*; Prague, pp88-95.

[7] Mikton J., 2008. *It is not about technology, it is about behaviour and norms,* DOI= http://beyonddigital.org/2008/12/21/it-is-not-about-technology-it-about-behavior-and-norms/ (last accessed 6 July 2015)

[8] Mikton J., 2014. The death of digital, DOI= http://beyonddigital.org/2014/05/22/the-death-of-digital/ (last accessed 6 July 2015)

[9] Oram D. and Headon M, 2002. *Conference paper. Avoiding Information Systems failure; culturally determined ethical approaches and their practical applications in the new economy*. International Scientific Conference, Kaunus, Lithuania, 18-19 April 2007.

[10] Patrignani N and Whitehouse D., 2013. From slow food to slow tech. *Proceedings of the IADIS multi-conference on ICT Society and Human Beings, ed P Kommers*; Prague, pp 160-164.

[11] Prensky M., 2001, Digital natives : digital immigrants, DOI= www.marcprensky.com/writing/Prensky Digital Natives, Digital Immigrant Part 1.pdf (last accessed 6 July 2015)

[12] Prensky M., 2012, From digital natives to digital wisdom. DOI= marcprensky.com/writing/Prensky-Intro_to_From_DN_to_DW.pdf (last accessed 6 July 2015)

[13] Preston C. and Payton M., 2015. Towards tomorrow's successful digital citizens A policy think tank report at the ITTE conference DOI= www.itte.org.uk (last accessed 6 July 2015)

Including Teaching Ethics into Pedagogy: Preparing Information Systems Students to Meet Global Challenges of Real Business Settings

Shalini Kesar
Associate Professor
CSIS Department
Southern Utah University
Cedar City, Utah, USA
+1-435-865-8029
Kesar@suu.edu

ABSTRACT
This paper's discusses five real business capstone projects that were designed to provide an educational experiential learning to include ethics and professionalism in the pedagogy. This class comprised senior undergraduate Computer Science (CS) and Information Systems (IS) students. The projects involved teamwork and lasted fifteen weeks. The aim of the capstone curriculum was to foster a teaching environment to: 1) include interdisciplinary partnership among university departments; 2) cultivate local industry alliances; 3) encourage students' analysis and synthesis of skills and knowledge in a real business setting project. The National Society for Experiential Education (NSEE) practices were used while developing the curriculum and their eight Guiding Principles of Ethical Practices are used to describe the initial findings and lesson learned.

Categories and Subject Descriptors
K.4.1 [**Computers and Society**]: Ethics

General Terms
Management, Documentation, Human Factors

Keywords
Capstone projects, experiential learning, soft skills. National society for experiential education, pedagogy.

1. INTRODUCTION
This paper is part of on an on-going research that discusses the importance of embedding professionalism and ethics component in an undergraduate senior capstone class projects pedagogy. This style of pedagogy will provide an experiential learning education environment, which will better prepare them to deal with challenges faced in today's technology related businesses.

The capstone projects are designed to give students the chance to apply the skills and knowledge they have acquired in the previous courses. Hence, capstone projects are considered to be a cumulative and integrating experience to be part of a real world problem in a classroom environment that facilitates critical reflection, enhances communication skills, interpersonal relationships, project planning and organization [1]. As pointed out by Kolb and Kolb [2] such projects allow "recursive spiral knowledge development".

The capstone classes described in this paper were designed to provide an educational experiential learning that fosters a teaching environment to: 1) include interdisciplinary partnership among university departments; 2) cultivate local industry alliances; 3) encourage students' analysis and synthesis of skills and knowledge in a real business setting project.

While developing the curriculum, various studies, which are discussed throughout the paper were taken into account. In addition, best standards and Guiding Principles of Ethical Practice by the National Society for Experiential Education (NSEE) were used. The NSEE Guiding Principles of Ethical Practices [3] are used to develop the pedagogy to teach ethics and professional as part of an experiential education. This paper also describes the anecdotal evidence and how instructor included ethics and professionalism in the undergraduate Computer Science (CS) and Information Systems (IS) course.

2. BACKGROUND
Lectures, case studies, guest lecture (including client's visit), and good source of articles (various codes of ethics) were part of the teaching material. The fifteen weeks of capstone projects were selected based on the criteria on how to enhance two main skills of the students: technical and core. Core skills included integration of critical thinking, improving effective communication, and being responsible and accountable while working as a team. In other words, core skills are soft skills and can be linked to experiential education whereas technical skills were to ensure students use and apply knowledge from previous

coursework to their real business project. The projects included different business context and were in the area of information literacy, database, forecasting and decision support, and web presence. This is a unique class because both IS and CS students were registered for this class. The class met once a week for three hours.

The National Society for Experiential Education (NSEE) founded in 1971, is an open society dedicated to mutual learning and support across a variety of roles and responsibilities represented in the field of experiential education. The main mission of NSEE is to foster the effective use of experience as an integral part of education and empower learners. They outline best practices and guidelines that can be used as a starting point for any field. The "Eight Principles of Good Practice for All Experiential Learning" outlines the conditions, steps and actions that are necessary for success. The instructor used these as a starting point together with various literature studies about incorporating ethics in a capstone class when developing curriculum for both CS and IS students.

3. CAPSTONE CLASS PROJECTS

3.1 Guiding Practices

This section begins by Eight Principles of Good Practice for All experiential learning activities suggested by NSEE, followed by how the instructor developed pedagogy to also include teaching ethics into the curriculum.

Intention: All parties must be clear from the outset why experience is the chosen approach to the learning that is to take place and to the knowledge that will be demonstrated, applied or result from it. Intention represents the purposefulness that enables experience to become knowledge and, as such, is deeper than the goals, objectives, and activities that define the experience

The most important point to keep in mind is the main intention of choice of capstone projects. The instructor had to ensure that both technical and core skills were met when creating learning approach. The instructor has been teaching the capstone class for seven years and as mentioned earlier the projects are different in context and all related to real life business settings. The projects specifically described in this paper are a few selected ones in different areas to demonstrate learning of knowledge of both hard and soft skills in the classroom. The main intention of developing such a curriculum was to: 1) to prepare students to have the ability to critically think about the ethical issues in real business world; 2) enhance awareness about such components along with their importance and challenges; 3) This kind of learning environment in the classroom will motivate students to incorporate the codes of ethics in their behavior; 4) raise the bar a bit to create a robust learning environment to include the importance of effective communication, codes of ethics, professionalism, team working, leadership skills. Hence, the course was structured to include ethical and professional aspects in the lectures, assignments and in the evaluation process.

Preparedness and Planning: Participants must ensure that they enter the experience with sufficient foundation to support a successful experience. They must also focus from the earliest stages of the experience/program on the identified intentions, adhering to them as goals, objectives and activities are defined. The resulting plan should include those intentions and be referred to on a regular basis by all parties. At the same time, it should be flexible enough to allow for adaptations as the experience unfolds.

The NSEE practices indicate that it is valuable to ensure that participants (in this case students) must ensure that they enter the experience with sufficient foundation to support a successful experience. While creating the lesson plans and projects, the instructor reviewed the technical skill sets, previous projects and work experiences of the senior CS and IS students. One of the methods to have better planning for a "successful" experiential education environment was to require student complete a questionnaire a few weeks prior to the spring semester when the class started. This helped the instructor to discuss with the business clients and identify a project. This allowed the instructor to adhere to the goals, objectives and activities identified as well to the learning outcome of the program in the department. Although the phases of the projects were outlined because they were real business cases, the instructor was aware of the importance of being flexible enough to allow for adaptations as the experience unfolds (see below for more details). Various definitions and underlying goals outlined in literature studies were kept in mind. Capstone course is a method of collective evaluation that assesses students' skills of previous learning and overall collegiate learning experience. In addition, Jervis and Hartley [4] point out that in order to effectively end a college career and begin a professional one, capstone courses should, for example, should include ways to facilitate a learning environment that helps them to transit from the academic to the professional world. While planning and preparing a capstone classes, it is also important to note that group work is an integral part of students' education As an example, this is manifested in the important role of teamwork in the Association for Computing Machinery (ACM) Curriculum [5], as well as in many study programs.

Authenticity: The experience must have a real world context and/or be useful and meaningful in reference to an applied setting or situation. This means that is should be designed in concert with those who will be affected by or use it, or in response to a real situation.

The main reason for including teaching ethics to both CS and IS capstone students was that their classroom experience should relate to a real world context which is useful and meaningful in every changing technology field. Most of them plan to work in technical fields where they will face global challenges that will not necessary be associated with technical issues. One of the projects was an ongoing project. Every year the project is from the same organization but is a different phase of the business. A few of the gradated students were invited to share their experience about various challenges they faced in real business settings. These students were employed by the same businesses after working on the capstone project in class. Students were encouraged to share some of ethical and professional dilemmas (this is discussed more in later sections).

Reflection: It is the element that transforms simple experience to a learning experience. For knowledge to be discovered and internalized the learner must test assumptions and hypotheses about the outcomes of decisions and actions taken, then weigh the outcomes against past learning and future implications. This reflective process is integral to all phases of experiential learning, from identifying intention and choosing the experience, to considering preconceptions and observing how they change as the experience unfolds. Reflection is also an essential tool for adjusting the experience and measuring outcomes.

In order for the student to work in a real business environment

experience, it is that they also discuss and reflect upon their experiences. According to NSEE practices, reflection is the element that transforms simple experience to a learning experience. In other words, reflection provides input for new hypotheses and knowledge based in documented experience, other strategies for observing progress against intentions and objectives should also be in place.

The reflective process both as part of discussion, presentations, and report writing was incorporated in curriculum. This allowed both the instructor and students to experience different phases of real businesses setting from identifying of global challenges and choosing the experience, to considering preconceptions and observing how they change as both individual and as a group as the experience unfolds.

Orientation and Training: For the full value of the experience to be accessible to both the learner and the learning facilitator(s), and to any involved organizational partners, it is essential that they be prepared with important background information about each other and about the context and environment in which the experience will operate. Once that baseline of knowledge is addressed, ongoing structured development opportunities should also be included to expand the learner's appreciation of the context and skill requirements of her/his work.

The NSEE points out the importance of organizational partnership. As mentioned earlier, the instructor encouraged guest lectures from different departments, including English, communication, and law. This provided background information and learning skill about technical writing, communication and presentation skill and relate the importance of theses skills in the context and working environment in which they will be exposed to after graduation.

Monitoring and Continuous Improvement: Any learning activity will be dynamic and changing, and the parties involved all bear responsibility for ensuring that the experience, as it is in process, continues to provide the richest learning possible, while affirming the learner. It is important that there be a feedback loop related to learning intentions and quality objectives and that the structure of the experience be sufficiently flexible to permit change in response to what that feedback suggests. While reflection provides input for new hypotheses and knowledge based in documented experience, other strategies for observing progress against intentions and objectives should also be in place. Monitoring and continuous improvement represent the formative evaluation tools.

In order to monitor student improvement, various individual and groups assignment, progress reports and online discussion forum was designed in the curriculum. Rubric system was used to allow the students to monitor their own progress (from grading point of view). This also allows the instructor to learn and as well receive feedback on the quality objectives and the structure to modify the curriculum of the next capstone class.

Assessment and Evaluation: Outcomes and processes should be systematically documented with regard to initial intentions and quality outcomes. Assessment is a means to develop and refine the specific learning goals and quality objectives identified during the planning stages of the experience, while evaluation provides comprehensive data about the experiential process as a whole and whether it has met the intentions which suggested it.

While planning the lessons, different types of assessments were deigned. The evaluation process of each assignment varied, for example, report writing, class presentation, public presentation. The design for assessment incorporated core skills evaluations. In the later section, the instructor discusses how outcomes, processes, and formal requirements were designed for this capstone class. These were systematically designed by keeping in mind the initial intentions.

Acknowledgment: Recognition of learning and impact occur throughout the experience by way of the reflective and monitoring processes and through reporting, documentation and sharing of accomplishments. All parties to the experience should be included in the recognition of progress and accomplishment. Culminating documentation and celebration of learning and impact help provide closure and sustainability to the experience.

The final practice refers to acknowledgement. It is important that student feel the sense of accomplishment. One of the requirements in the group report relates students acknowledge individual and group progress. Students were acknowledged during presentations and during their public presentations. A section of the class presentation included students acknowledging their team members and clients. It also included students providing a brief talk on what technical and softs skills they learned from each other in a team.

3.2 Projects

This section briefly describes five projects that were included to create a real business setting capstone class. Although the project overlapped and some core content taught in the classroom, for the purpose of this paper, they have been identified and labeled as information literacy and web presence, innovation project, database, forecasting and decision support and in-house.

Information literacy and web presence project: This project involved creating a web presence for a local business and developing a training and maintenance manual. The local business used technology only for their inventory. Students were required to use different tools, liaise with employees, create a feasibility study, training program that included aspects of maintaining the website. Liaison with clients was in class and on site. The students had to be careful not to overwhelm the client and their employees with technological jargon while developing a web presence for the business.

Innovation project: This project involved developing a new platform via video games for high school education. It required students conducting research that took into account feasibility (technical, financial, and practical); content of the subject; presenting different platforms, acquiring software licenses, liaison with technical staff on campus for authorized access to virtual machine to download the software; working in a team to outline deliverables for the semester; allocating responsibilities based on the skills of the team members. The team members included

majority IS student and a few CS students. This perhaps was the most diverse group of students including international.

Database Project: This capstone project was for the railroad industry. This was an on-going project that has been incorporated in the class for the last four years. This particular phase of the project dealt with CS students working on developing a database and a web application that would connect the different divisions of the railroad industry thereby improving safety and productivity. Students conducted research on different programming languages, created a checklist of the rules and regulations of the company, allocated different responsibilities for team members, ensured weekly reports were posted (tasks, challenges, and possible solutions). Since students were authorized to access sensitive data, a non-disclosure agreement was involved.

In-House Project: Another group (both CS and IS) worked on developing a database for the lost and found division for the University post office. This in-house involved not only the postal service but also the Police and Safety division of the university. Lost items valued more than a certain dollar amount were kept in the Police and safety division. The main goal of the group was to develop a database that would be effective and easy to maintain. It was important that the University's policies were kept in mind while planning, conducting research and feasibility study. Communication and learning a new programming language were perhaps the two most important aspects in the course of this project.

4. DISCUSSION

The various projects described in this paper were related to real business settings. The discussion in this section uses the eight NSEE "Guiding Principles of Ethical Practices" [6] as a framework It follows two main themes: one, it relates to the specification of what the instructor incorporated in the curriculum and, second, various lessons learned from both instructor and the students' viewpoints. In addition, it highlights the unique challenges of developing as well as teaching such a course.

Principle One: Experiential educators uphold the principles of engaged education and democratic societies, the pursuit of truth, and the freedom of students to express their viewpoints, engage in critical thinking, and develop habits of reflection and civil discourse, listening and learning from those whose experiences and values differ from their own.

Prior to the first day of class, all students send the agenda and overview of the business projects. On the first day of class, the projects were discussed by both the instructor and the clients. The students were provided an opportunity to ask questions. It was interesting to see how a varied point of view of the students. Some students were more concerned with the timeline and the usefulness of such projects that required working with business. Various professional practices such as Code of Ethics, Code of Conduct were explained and case studies were discussed in class to highlight their importance.

By the end of fifteen weeks, the same students who questioned the intention of the purposefulness and the experience gained in the classroom were appreciative of the exposure they received to a real business setting.

Principle Two: Experiential educators use recognized, quality standards and practices in the placement and supervision of students engaged in field-based learning experiences and in the creation and maintenance of ethical partnerships with the communities and organizations that host and support these students, maintaining privacy, confidentiality and reciprocity throughout. The goals of course projects are to prepare students for the working life, making them familiar with the work place by practicing their skills on real-world business setting. This capstone class required students to visit the sites and then report on their finding and share their experience in the classroom. They were required to update their weekly report on Google drive as a group. There were to include their weekly goal as a team individual task, challenge encountered, and how did they overcome the challenges as both a group and as individuals. Depending on the project, site visit was appreciated and a new learning experience for the students. This is because students were interacting with other departments and professionals in the business. This allowed the students to not only appreciate today's changing business environment but also the importance of soft skills to be successful. One of the successful outcomes was that at least two students were hired from capstone class each year. The capstone experience enabled these students to develop skills that they can use immediately to contribute to the businesses that hired them. A skill set of integration of technology with a multi-disciplinary component [8, 9, 10].

Principle Three: Experiential educators recognize the depth of responsibility in teaching and modeling the values, skills, and relationships that foster a spirit of inquiry and fairness without discrimination or disempowerment.

To ensure curriculum is designed to foster a experiential learning education, the instructor attended a NSEE workshop and training where she received certification. The updated information provides a support system to modify and update curriculum as needed. To include teaching ethics, the instructor also attended workshops designed by the ACM.

Principle Four: Experiential educators are informed and guided by a body of knowledge, research and pedagogical practices recognized by and specific to the field of experiential education, including reflection, self-authorship, assessment and evaluation, civic engagement, and the development of personal and social responsibility.

As mentioned earlier, the instructor has received certification by the NSEE for training in pedagogical practices in the different fields. In addition, presentation in conference creates a sharing platform on the importance of modifying teaching pedagogy to include needed topics such as ethics, professionalism, and soft skills.

Principle Five: Experiential educators are committed to excellence through active scholarship, assessment and instruction, and the creation of shared knowledge and understanding through affiliation with networks and organizations that advance experiential learning.

Regardless of the type of the capstone project, the rubric for assessments, presentation and group report where students reflected on their progress and accomplishment was divided equally. As stated by the NSEE, reflection is also an essential tool for adjusting the experience and measuring outcomes.

Students were required to present progress report three times in the semester. As a group decided, they made a decision to choose

the date of presentation. The outline of the requirements, a few rules of presentation, grading system m were posted electronically in the beginning of the semester. Students enjoyed the freedom of choosing the time frame when they wanted to present the progress. This also provided a means to create a sense of responsibility, accountability, and professionalism as a team. Literatures studies highlight that such skills are crucial when preparing students when facing global challenges in the work field. In their paper, Leidig and Lange [11] highlight lesson learned from one hundred projects over the course of ten years. Although their focus was on community-based non-profit organizations, they provide useful insight about information systems capstone.

Principle Six: Experiential educators create informed learning contexts that foster student growth and actualization of potential, achieve academic and civic goals, and reflect excellence in curriculum design and quality.

Quality and design of the curriculum of the capstone class was validated by the same students were hired by the businesses. It could be argued that the clients had found an opportunity to develop the skills in the classroom catering for their businesses. However, most of the students hired were hired mostly for their soft skills and ability to work in an ever changing environment with individual and group challenges.

Some students resisted the requirements of the course, such as presenting to a public audience and peer evaluation on the final exam day. However, it was interesting to note that it was the those same students who actually felt enriched and said it added "value" to their experience. Discussion of various ethical dilemma with the projects (without the client presence) and how to overcome such challenges by relating to the lecture notes of case scenarios provided a good foundation to experience real life business setting hypotheses about the outcomes of decisions. Consequently the actions taken, then weigh the outcomes against past learning and future implications.

Principle Seven: Experiential educators are aware of and sensitive to recognized legal, ethical and professional issues germane to the field of experiential education and act in accordance with established guidelines to ensure appropriate practice.

Recognizing the sensitive of the real business projects where students have access to sensitive and confidential data, the instructor provided material about Codes of Conduct and Codes of Ethics and Professionalism by the ACM. It was also important that students are aware of the legal ramifications. Although the project were designed for fifteen weeks, students were required to sign legal documents such as Non Disclosure Document (where applied). One of the main beneficial was to enhance awareness of students about the complex nature of project and involvement of different division not necessarily from technology.

5. CONCLUSION

Overall students appreciate the pedagogy style of teaching and believed that that experiencing core skills added value to the class. Core skills such as integration of critical thinking, improving effective communication, and being responsible and accountable is very critical in today's technology business. The intention of choosing projects of different business context provide an exposure to the class of the different complexities and business settings. The instructor role over the span of fifteen weeks changed to make adjustments in teaching while overcoming business challenges. The instructor's role (the author) in these projects changed from that of an instructor to the role of a facilitator of learning and coordinator of experiential learning environment.

It was interesting that this pedagogical approach initially did not spark interest in the students. However, as the weeks passed, lectures and real case study scenarios related to the student's project made it interesting to hear about different viewpoints and reactions to the same situations. Majority echoed the complexity of their project experience was not due to technological issues but was more to do with "people" and "ethical" aspects.

To conclude, pedagogy for CS and IS capstone class should be an experiential educational that includes core skills such teaching ethics and professionalism. This will prepare the students to face the global challenges in today's technology based businesses.

5.1 Limitation

Any learning activity will be dynamic and changing, and the parties involved all bear responsibility for ensuring that the experience, as it is in process, continues to provide the richest learning possible, while affirming the learner. This paper provides reflects on the modified pedagogy to including important issues such as ethics to prepare student for the global challenges. Having said that this research is on-going and is limited to initial findings.

5.2 Future Direction

The instructor aims to continue this research to collect data by questionnaire. It also provides an in-depth analysis of the feedback provided via Google forms.

6. REFERENCES

[1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. ACM Trans. Program. Lang. Syst. 15, 5 (Nov. 1993), 795-825. DOI= http://doi.acm.org/10.1145/161468.16147.

[2] Kolb Kolb, A. & Kolb, D. (2003). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education*.

[3] The National Society (2009). Guiding Principals of Ethical Practices. DOI= http://www.nsee.org [4] Jervis, K. J., and Hartley, C. A. (2005). Learning to design and teach an accounting capstone. Issues in Accounting Education, 20 (4), 311–339.

[5] Computing Curricula, 2005) Computing Curricula, T. J. T. F. for. (2005, September). *Computing curricula 2005*. DOI: Http://www.acm.org/education/ curric_vols/CC2005-March06Final.pdf

[6] The National Society (2009). Guiding Principals of Ethical Practices. DOI= http://www.nsee.org/guiding-principles.

[8] Keller, S., Chan, C., & Parker, C. (2012). Generic skills: do capstone courses deliver?. In *HERDSA 2010. Refereed papers from the 33rd HERDSA Annual International Conference*, (pp. 383-393). HERDSA. DOI: March 15, 2013 from: http://dro.deakin.edu.au/view/DU:30030404

[9] Kumar, A., Baker, K., & Ahmed, I. (2004). Designing a Capstone Course for Information Systems: Challenges Faced and Lessons Learned. *Issues in Information Systems V* (1), 173-179.

[10] Jones, J. Empowering student in the information systems capstone, Issues in Information Systems Volume 15, Issue II, pp. 311-320, 2014

[11] Leidig, P. M. and Lange, D. K. (2012). Lessons Learned From A Decade Of Using Community-Based Non-Profit Organizations In Information Systems Capstone Projects. Proceedings of the Information Systems Educators Conference ISSN, 1435.

Musings on Misconduct: A Practitioner Reflection on the Ethical Investigation of Plagiarism within Programming Modules

Michael James Heron
Robert Gordon University
Aberdeen
Scotland
m.j.heron1@rgu.ac.uk

Pauline Belford
Dundee and Angus College
Arbroath
Scotland
pauline.belford@gmail.com

ABSTRACT
Tools for algorithmically detecting plagiarism have become very popular, but none of these tools offers an effective and reliable way to identify plagiarism within academic software development. As a result, the identification of plagiarism within programming submissions remains an issue of academic judgment. The number of submissions that come in to a large programming class can frustrate the ability to fully investigate each submission for conformance with academic norms of attribution. It is necessary for academics to investigate misconduct, but time and logistical considerations likely make it difficult, if not impossible, to ensure full coverage of all solutions. In such cases, a subset of submissions may be analyzed, and these are often the submissions that have most readily come to mind as containing suspect elements. In this paper, the authors discuss some of the issues with regards to identifying plagiarism within programming modules, and the ethical issues that these raise. The paper concludes with some personal reflections on how best to deal with the complexities so as to ensure fairer treatment for students and fairer coverage of submissions.

Categories and Subject Descriptors
K.7.4 [**Professional Ethics**]: Codes of ethics; Codes of good practice; Ethical dilemmas.

General Terms
Security; Human Factors; Legal Aspects

Keywords
Plagiarism; Programming; Teaching; Ethics; Morality; Attribution; Academic Misconduct; Education

1. INTRODUCTION
As a necessary part of evaluating student work, teaching professionals must assess its originality and conformance with institutional rules of attribution. Plagiarism is an unfortunate occurrence within student work, and thankfully as best as can be ascertained still a minority phenomenon. Automated tools such as *turnitin* [3] have allowed for much plagiarism to be automatically identified and the original sources to be located, and the provision of such tools to students for self-assessment even discourages attempts to submit problematic work in the first place [7][8].

Within software engineering, and specifically the field of programming, dealing with plagiarism is much more difficult.

Standard tools such as *turnitin* do not offer facilities for checking the originality of software solutions. While tools such as MOSS [1][5] exist as an attempt to detect similarity in code there are elements that are unique to software development that limit the utility of such automated routines. In the end, it is down to an academic to analyze the code, and usually within the tight time constraints implied by assessment boards and other formal duties. As such, not every submission will receive the same amount of critical attention. Those submissions that are most suspect will receive the greatest amount of effort with regards to investigation. Given the relatively fine-balanced mesh of issues that determine the quality and originality of programming code, the designation of a submission as suspect is often a matter of academic judgment. This presents numerous ethical issues for those who must ensure the integrity of assessments.

In this paper, the authors reflect upon these ethical issues as a professional educator. This paper does not offer a better system for dealing with potential plagiarism in software development modules – there exists, at this time, no obvious alternative to that of relying on the academic judgment of subject matter experts. However, this paper intends for the discussion to help illuminate some of the important considerations that such a state of affairs raises. The authors hope that fuller understanding of the problem helps ensure that students are given the fairest possible consideration when such incidents are investigated.

2. GOOD PRACTISE AS PLAGIARISM
In many ways, it is difficult to truly render a verdict of 'plagiarism' in software development without first invalidating many of the fundamental lessons we attempt to impart to students regarding how programming works in the real world. Some of these issues are already well understood – we work within a medium where vocabulary and syntactic construction of the simplest elements is ritualistic to the point of incantation. Programmers cannot simply extemporize or add lyrical flourishes to an argument to underscore a clever point. We must work within the constraints of the programming language's grammar. Programmers, working on the underlying bones of a program, are limited to three key forms of expression – linear, loop and selection. There are only so many ways to write a for loop or an

Ethicomp '10, September 7–9, 2015, Leicester, United Kingdom.

if statement, and certain standard cultural conventions regarding names for variables and layout of code have taken deep root in both professional and educational instruction. Consider for example the names **i** and **j** as counter variables – while we may all acknowledge how ineffective these names are, as part of a common vocabulary of programming they are hard to ignore.

Not only is the structure of a program often impossible for us to meaningfully alter, but so too often are the unspoken assumptions that we absorb via osmosis through exposure to a larger, established community of practice. As part of that community of practice, we absorb understanding – first informal and then gradually formalized into *standards* – about how code should be written. We discourage experimentation with format, layout and even the name of variables because such things represent **bad practice** [10]. In this way, we sanctify certain plagiaristic practices, turning them from vice to virtue.

Taking a step back from the raw bones of individual statements and structures, a common component of programming courses tends to be some kind of formal instruction in the topic of algorithms. We teach students to understand Big O notation and explain the relative merits of bubble versus quick versus merge sorts. During these discussions, we underscore why we use algorithms. Rather than attempting to reinvent the wheel, we rely on tried and tested solutions to complicated problems because these tend to be more reliable than more original and creative solutions. If the syntax and conventions of programming limit the vocabulary of a programmer, algorithms work to constrain computational creativity.

When students gain a little more appreciation of the way that object orientation works, we may introduce them to the wider world of design patterns, explaining loftily that design patterns are to objects as algorithms are to processes. We go through the classic architectural relationships implied by the most widely used patterns, showing scenarios in which they can be used and encouraging students to consider where to apply them in their own code. 'They may not be the best solutions, but they're good solutions – battle tested solutions', we say. We grade our students abilities to both interpret design patterns in the work of others, and apply them to their own projects. In this way, we place shackles even on the way in which objects within a program are expected to communicate.

We stress the value of **reusability**, often honoured more in the breach than in the observance in the real world, and nod approvingly when reusable code is produced by our students. We favourably grade that code which tightly conforms to general design principles such as encapsulation, remarking with good grace that the submission offers good scope for reuse in other programs. We make a big point of arguing the importance of **maintainability**, stressing that most of software development is in maintenance and that good programs can not only be reused in their current form but refactored to work in other areas too. We encourage, certainly in later years, the use of external libraries to do the heavy lifting in problem domains where students cannot reasonably be expected to 'roll their own' solutions.

We may do all of these things, or some of these things, or none of these things – every academic has their own set of developer battle-scars that experience has cut into their skins, and the way in which particular messages will be emphasised will be a function of this. However, the reality of what it means to develop software, and the lessons we teach students about the development of code, is often starkly incompatible with how we **treat** code which honours the lessons that we have taught. Students often do not realise that what they are doing could be construed as plagiarism because in many ways it's just following the advice they've been given about how code should be written.

3. PLAGIARISM IN PROGRAMMING

It is here where academic judgment becomes an important, and ethically troublesome tool. It is the responsibility of an academic to assess a piece of work in its entirety and form a judgment as to the level of originality shown in a submission. Depending on how strict we wish to be about definitions, it is reasonable to argue that all programming is plagiarism to one degree or another. A University of which I am aware and which will not be named, once formalised its institution-wide plagiarism policy with a requirement that students attribute every single thing that they didn't write themselves. This was held to be the case even if it came from their own lecturer's slides and that failure to do so would be considered a breach of academic conduct.

This policy was constructed without reference to the School of Computing, who would have pointed out that this mean that every single program produced by every single student in every single module would require every single line of code to reference some standard text in programming. This purely as a consequence of the limitations of grammar imposed upon practitioners. Even firm policies regarding attribution of 'anything not in a lecturer's slides' are inconsistently applied – why for example do we need an attribution from Stack Overflow, but not one for using a whole set of Javascript tools such as jQuery? Why do we need to cite a string tokenization tool we grabbed from Unity Answers, but nobody needs us to cite an Abstract Factory? Why is it okay to use a graphical asset from the Unity Store, but not a tutorial on the Unity website? The edge cases here are many because of the need to find the right point in the spectrum between 'attribute every line' and 'attribute nothing'.

The problem is further complicated by the many ways in which plagiarism might be reflected within software code, and the degree to which software development is an incremental process. A program of ten thousand lines may have an incredible structural dependency on a handful of objects at the core of a vast class relationship. Classes may be incredibly light at the core but become much denser the more specialised they become. A piece of code may be complex in its functionality but marginal in its effect, and vice versa. Unlike in an essay where each word should plant a step in one ongoing journey (ideally), a computer program is more like the schematic for a complicated machine through which information will flow in unpredictable ways.

Plagiarism then might be in individual lines of code, in the collection of code into functions or objects, or in the relationship between classes and objects. It might be in the way in which an Application Programming Interface (API) is exposed, and arguments have been made that this should even extend to the order and type of parameters sent into functions [9]. It can also be easily masked by students looking to mislead an academic or just by those who didn't realise that what they were doing strayed into plagiarism at any point. Within my own university, we focus on attribution as the differentiator between 'good software development' and 'plagiarism', but that presupposes that students are aware that there is a need for attribution at all. In the process of diligent software development, a student may refactor a piece of code taken from elsewhere to add in features, remove unnecessary complications, or simply make it consistent with the context in which it is placed. Thus, while they are benefitting

from a solution they have found elsewhere, they gradually smudge over its original 'alien' conventions and bring them into line with the conventions they themselves use. The cracks between the two sets of code are plastered over, and if done well there will be no sign that there was ever a crack there to begin with. Thus, the plagiarism becomes, through adaptation, completely invisible. That does not mean it was never there, but the skill and knowledge required to both understand the code and refactor it for consistency is in itself a very valuable programming skill. However, attribution would still be necessary to acknowledge the intellectual debt that the submission owes to the original author, even if there is no trace of the original code left. In my experience students rarely go to the effort of consciously (or indeed, blissfully unconsciously) covering their tracks – they just don't realise that what they are doing constitutes a need for attribution [4].

In many cases, code might not simply be taken from an online or offline source but written collaboratively with other students. In such collaborations, it is rare that the effort is invested equally amongst all participants, and often the stronger students give more assistance than we might desire to their weaker colleagues. However, one of the key benefits that comes from a formal educational experience is the social context in which study is placed. We expect students to discuss their work with each other, plan out solutions, confer on tricky sections, and so on. All of this is useful team-building and group-work – elements that we often work to explicitly stress within core elements of a curriculum. However, we still do expect when work is submitted that it is meaningfully distinct for each individual. Within the constraints of software development though, this can be difficult and students often lack the skills required to make a meaningful judgment on what constitutes distinctly original work. We expect this work to be different in more than just a few variable names or function names, but no matter how we may stress this we are still operating in an environment where the skills to make that judgment may be lacking.

Thus, we see multiple submissions of what is essentially the same code, with only minor surface details changed. Here then is plagiarism which is merely a virtue taken too far into vice – there is often no intent to deceive, or the effort to hide the source of code would be performed more diligently. Students too in these circumstances often confuse the difficulty they had with the work with the academic's likely judgment on individual effort. They may not realise that the fact it took four hours to write a loop doesn't mean that it looks like four hours of effort to the person grading it.

In the experiences I have had with student counts of plagiarism, of which there have been many, only a very small fraction of them have left me feeling that there was a genuine attempt to obtain through deceit credit for work that they had not done. Instead, it tends to be one of the following:

1. Believing they were more responsible for the code that they submitted than a dispassionate review of the contribution would reasonably conclude.
2. Being unaware of the need for attribution in an environment where reuse, generalisation and reliance on external libraries is permitted, and even encouraged.
3. Not appreciating the line between healthy collaboration with colleagues and plagiarising from class-mates.
4. Not fully understanding the expected attributional difference between exemplar material written by their lecturer and provided within the context of a course and those external resources which may be mentioned as 'further reading'.

Clear communication of these issues helps, but it presupposes again that students believe that the communication applies to them, and that they'll remember it when it comes time to submit the work they have done. In the latter case, the stress of deadlines and the worry over the degree to which a submission meets a coursework brief can be distracting enough that attribution may simply be a distant thought in a head already full to bursting.

That is not to say that we should not take a strong position on work that is judged to have been plagiarised, but instead to outline some of the complexities that come with ruling that a piece of work is plagiarised at all. Reasonable people can disagree on the extent to which a piece of programming code represents original work, acceptable modification of the work of others, or outright plagiarism. When it's difficult for subject matter experts to agree on all the details, it is especially difficult for students who lack the training and understanding of the wider context that experience provides in slow, gradual accumulation.

4. IDENTIFICATION OF PLAGIARISM

Within the process of identifying plagiarism, we must resort to academic judgment to determine when a submission has fallen over the line between 'good practice' and 'intentional or unintentional deceit'. The number of submissions that we must routinely analyse along with the intricate complexities of each individual submission mean that we can only ever truly, feasibly, investigate a proportion of these. Tools for automating detection are, for software code, lacking in the sophistication to pick up on anything other than the most overt use of external sources. Thus, we must choose a sample only. In the next section, this paper will discuss some of the ethical implications of this selective analysis.

Informal suspicions may be initially raised in a number of ways. This paper will outline these in turn before moving on to the ways in which the original source for code may be located. Of a necessity, we will not be too specific about the full range of ways in which plagiarism may be identified as the task is already difficult enough without adding additional elements of challenge. Much of the process must be shrouded behind a kind of 'security through obscurity' model.

One of the things that happens for the majority of students within software engineering degrees is that a faculty builds, from the ground up, an understanding of programming. We lay the foundations of their understanding, choose the examples, and structure the assessments. Within a faculty we might cover a broad range of skills and styles, but there is often a link between early and late parts of the curriculum embedded in a single individual. The one that teaches first year programming may also be the one that teaches second year programming. If that is not the case, the necessity of understanding the context of a student's overall experience of a topic means that lecturers will be aware of what is done in the pre and co-requisite modules that describe their own course's academic context.

We also sample the code that students write during practical exercises, often seeing the evolution of coursework as it is moulded from rough sketch to polished artefact. We likely have a hand in that evolution, offering suggestions here, corrections there, and an occasional helpful hand in tracking down misbehaving subsystems. This kind of ongoing familiarity means

that we can see the way each individual student writes code, and we can see the degree to which it harmonizes with the way in which we've been teaching the topic. Everyone has their own particular quirks when teaching programming – some favour associative arrays, some prefer arrays of objects. Some prefer the strict architecture of a formally designed class model. Others prefer a looser, ad hoc arrangement of code. Some favour certain design patterns, others make use of language features that obviate their requirement. It is impossible to be a programmer without picking up some developmental quirks that represent the best solutions that have evolved from long, hard experience. On top of these are entirely ornamental quirks such as the way in which variables are named, or the use of camelCase versus underscores_in_names.

Within the courses we teach, many of these quirks will be communicated to students in the form of exemplar code, lecture content, or the occasional aside delivered as part of an informal discussion. These quirks in turn make their way into student submissions to a greater or lesser degree. My own propensity to use the word 'bing' as a temporary variable name has mentally mutilated any number of my students. I apologise if anyone reading this has had to deal with the consequences. As a result, the code that is produced by the students will tend to take on a signature that is similar to the one demonstrated by their instructors, and ongoing familiarity with what they are doing within the labs will make that signature known to their lecturers. It's something like our own personal accent – it doesn't **uniquely** identify us, but it will certainly be something people use to differentiate.

Thus, when code is submitted that doesn't conform to the signature we are expecting, it creates the first sense that something may be wrong with the code that is provided. It might be written with unusual formatting, strange variable names, or even in a structure that is entirely inconsistent with what we may have taught. In a module on using HTML5, we may find jQuery being used rather than the canvas we had been discussing; in a module on PHP, we may find that an old, clunky version of the mysql interface functions were used rather than the up to date mysqli libraries we had advocated. Such things don't necessarily mean that a student has taken their submission from another source, but do raise the suspicion that something unusual has been going on.

Rarely is it the case that such incidents spread throughout the entirety of a submission – what is more common is the discordant tone of two different styles clashing with each other. We are expecting to hear one accent, and suddenly in the middle of a sentence it switches to another – it has exactly that kind of jarring impact when we encounter it, and it too is a sign that something unusual has happened with a submission.

Sometimes it's not an especially jarring accent change, but instead a remarkable quality change – if the majority of a program is of dubious quality, but it surrounds a core that is beautifully written and designed, then we must treat the submission with suspicion. Similarly, if there is a beautifully designed program that just happens to be of dubious quality in those aspects of the brief that were least likely to be present in an online forum, we must consider the possibility of some form of plagiarism. Often, when writing assessments, a lecturer might use a standard 'stock exercise' that is well understood and easily communicated. In such occasions, a common tactic to dissuade students from using the first online solution they can find is to modify the specifics of the exercise to include aspects that are unusual. Thus, students may find the core of the solution but be left with the task of bashing at it until it does what the lecturer has thrown into the brief as a complicating factor. It is at these points of stress that we can often see the suggestion of some kind of code adaption.

Sometimes the suspicious aspect comes in with a student who dramatically over-accomplishes in functional requirements that were never part of the brief, but under-accomplishes in requirements that were. Such unusual prioritization of development time is suggestive that at least some of the submission may have come from a template which did not precisely map on to the requirements as outlined or emphasized.

As a result of familiarity with students during ongoing instruction, we also build up a reasonably good mental profile of which students are especially capable, and which require our additional support. Those students most needing support are usually also those that produce code with which we are the most familiar as we spend a greater proportion of our time working our way through it with them. When a student with whom we have been spending much of our time suddenly submits a piece of work that we strongly suspect is beyond their demonstrated capabilities, then that flags up our interest.

Suspicion however is not sufficient for conviction, and having had their attention drawn to a piece of work a lecturer must ascertain whether their suspicions are grounded. This is often a straightforward matter of finding an especially distinctive piece of code and throwing it into Google. A distinctive piece of code is usually one that is sufficiently complex that its presence acts as a fingerprint for some other project elsewhere on the internet. Students may, as a result of submitting such code, change variable names, the order of invocation of certain statements, or the values associated with variables. However, other pieces of code are less pliable – especially if they implement formulae or make heavy use of structural systems of the host language. In those cases where Google can't throw any light on the matter, the search must move on to other sources such as GitHub or other code archival sites.
If that doesn't work, it's possible to attack the problem from the other direction and execute a search for what you'd look for if you were trying to find a solution to your own coursework exercise. Sometimes the code is taken from a particularly obscure location, but it is rare that it takes too long to track down the original source of the code. In those cases where the code does not seem to exist, then it's necessary to consider the other plausible routes for the source.

More and more commonly these days, we must consider the source of a submission as being that of an essay mill [2][6]. Sadly, in such events where the providence of code may not be identified with online checking we must resort to whatever internal mechanisms we may have available to ascertain student understanding of their own submissions. My own preferred route is through a mini-viva, in which students are asked to explain how their submission works, and to outline the process through which they may have developed it. On occasion, such a mini-viva results in a student giving a considered and confident explanation that resolves any lingering uncertainty about the authorship of the work. Often too, the viva reveals a lack of understanding that likewise settles the issue in the other direction.

5. THE ETHICAL IMPLICATIONS OF INVESTIGATING PLAGIARISM

Having outlined the ways in which plagiarism may manifest itself within programming submissions, and discussed some of the ways

in which plagiarism may be detected by academics, we must turn to the ethical implications that are raised by such methods. If we are to truly treat students fairly, we must be aware of the troubling aspects of a process like this and examine where we can make systemic and procedural improvements to alleviate some of the issues.

First we must address the nature of student expectation – as discussed above, my own experience is that by and large students simply do not believe they are doing anything wrong. No matter how we may codify submission requirements, or inculcate a need to attribute, students often have a difficulty in seeing where the line between 'good software engineering' and 'academic misconduct' lies. There is a sector-wide inconsistency in how we teach software engineering principles, and how we treat students who adhere to principles of re-use. It is not that, as a sector, we do not communicate the importance of attribution – it is that students, as a general grouping, are often unaware of what should be attributed. The fact that there is rarely any obvious intention to deceive suggests one of two possibilities. The first is that students simply don't take any pride in their cheating, or have very low expectations of their lecturers. I don't believe, generally speaking, this to be true – it is not that there is little effort to obfuscate code, it is that there is often **no** effort to obfuscate. To the authors of this paper, that argues for the second interpretation – that such submissions are evidence of a lack of understanding, driven in part by the uneasy tension between plagiarism and sensible software engineering.

Solutions to such problems must stem simply beyond lectures on plagiarism and academic misconduct – students can easily give a word for word definition of plagiarism, and their responsibilities in that regard. It is not in the communication of the rules that we find problems, but rather in the interpretation.

With this in mind, we must always, first and foremost, look to whether we are properly contextualising the lessons of software engineering within their academic context. We should include discussions of what authorship means within software engineering and the day to day importance that attribution and sourcing plays in developing computer programs. We must also ensure that students take the necessary time to reflect upon the implications of their own submissions. Requiring students to formally acknowledge that the code they have submitted is entirely their own work, perhaps via a formal cover sheet, gives an opportunity for pause before uploading or sending the work. That pause might be what's needed to make them think 'Oh, I'll just put that attribution in, just in case'.

When assessing a submission for discords and disharmony as discussed above, we must also be mindful of the fact that in some cases we may have had only a small impact on the development of a student's personal signature. Students may have learned how to code outside our classes, and may indeed have arrived in the classroom with their own largely fully formed signature. If a signature is comprised of bad practice, our job may be to break it down and rebuild it in a better form. We must be mindful that any disharmonious elements in a code submission may be as a result not of external parties influencing a submission, but instead our own influence impacting on an already existing coding style. We must be careful to assess all submissions on their own merits, in the context of a student's own academic journey. Failing to do so could potentially subject a student to a harrowing hearing on academic misconduct where their own lack of a confident voice is used as evidence against them.

Similarly, as part of regular lab exposure to students we may find our own code making its way gradually into a submission as we explain how to address a problem or deal with a persistent error. Such *ad hoc* instruction tends to make the rounds amongst other students within that social circle, as it is usually perceived to be a 'lecturer approved' solution. It's important that as we provide such additional support to students that we realize that it is likely to be repeated in other submissions as the work is discussed and analyzed. If we are forgetful of what we have told our students, this can look very much like a whole group of students copying each other. In reality it is a piece of ad hoc support that we ourselves provided that has been traded around a class in response to others having the same problem. In such cases, we must be careful of alleging any misconduct at all – in real terms, there is little difference between students using our lecture notes and using the code that we may have provided, in passing, as part of private classroom discussions. Consider if we might, under other circumstances, have simply written the code out on a whiteboard for the class rather than doled it out to one or two individuals in the course of class discussions. In such cases, how do we even attribute authorship when it was not actually the student who was the source of the code?

When identifying submissions as being suspect or including elements worthy of deeper investigation, we must consider whether or not our own investigation has a bias built into it. We are unlikely to indifferently find submissions where the plagiarism has been done well - when students have managed to successfully marry disparate elements into a coherent and harmonious whole. When we identify work that seems to be sourced from elsewhere, we must be mindful to not simply focus on the low-hanging fruit else we run the risk of punishing those who try the least to obfuscate a submission. In addition to picking up on courseworks that are problematic, I advocate subjecting an additional random sampling of all submissions to an in-depth investigation even where there is no suspicion of wrong-doing. While such investigation rarely yields results, it has on occasion uncovered an especially clever piece of academic misconduct that would otherwise have gone unchallenged. In addition, it ensures that it is not only academic suspicion that leads to investigation – while such judgment is vital in uncovering plagiarism like this, it is also difficult to disassociate from the context of a student cohort. We cannot be sure that we are not letting personal likes or dislikes have influence on the investigation process. We cannot know how widespread plagiarism is within our modules – we can only say how often we notice it. By ensuring we sample beyond the obvious suspects, we can build our own confidence that the work we would otherwise have passed without comment is academically sound.

We must also be careful in ensuring that we are fairly representative of how we search out plagiarism. As discussed above, it may be extremely difficult to source code that comes from an essay mill, whereas a standard online tutorial may take only a few minutes of searching. This creates something of a class divide in investigating plagiarism, where those who can afford to buy 'off the shelf' solutions to class exercises are simultaneously inoculated against proper academic inquiry into the providence of code. When searching out the sources of work, we should look not only for the source of code, but also for incidences in which our course-works themselves have been floated online. Often, a search for a few indicative phrases from our own coursework briefs will reveal a request for a solution on an essay mill site, and this can be enough to raise real concerns regarding the authorship of submissions. An in-depth

investigation of a student submission involves taking in a number of sources, and it would be unethical to do so without considering what financial solvency may permit in terms of covering the true source of code authorship.

In the sourcing process too, we must be mindful of the fact that there may be several sources which have been synthesized into a single submission – it's unusual that only one source is ever the single canonical reference point for all incidences of plagiarism. It's not enough to simply find a bit of code and say 'gotcha'. It's necessary to forensically outline the source of code statements and consider whether the welding of disparate elements may reflect sufficient mastery of the topic in and of itself to be worth credit. For my own purposes, when I suspect plagiarism I will go through each line of a submission and comment out those that come from an external source. Where adjustments have been made, such as changing the name or value of variables, I will comment those changes too. The result is a review of the code that allows for me to specifically reference lines of code and link them back to their original source. That which is left is, as best I can tell, the student's original contribution to the work. On occasion, when mitigating factors have been taken into account, that original contribution can turn out to be sufficient to pass a module. Whether that is an appropriate outcome is something that must be assessed on a case by case basis, but this forensic deconstruction is a process that both serves to solidify an argument for academic misconduct as well as more effectively frame the student's own contribution to the work.

This forensic examination of the code can serve as a valuable part of a formal or informal viva on the providence of a submission. However, here we must be careful – academic regulations may not permit a viva to be used as an additional, unannounced format of assessment. Often as part of an academic misconduct hearing there will be some viva element in which students may be asked to explain their code, but this is different to simply getting people in to ask about what they did. The possibility of later examination via viva should be announced in course books and module descriptors. It would be unethical to assess based on hidden criterion within a course, and likewise unethical to offer no guidance as to who is likely to be selected to perform. Linked to this is an issue of stigma if the only people asked to present their work orally are those who have likely plagiarized – in such cases, the invitation alone is enough to overlay a degree of suspicion amongst students. Thus, if vivas are to be conducted they should include a random sampling of students who are under no suspicion of plagiarism. Not only does this mitigate the stigmata issue, it also ensures that there is a control group against whom performance can be calibrated. The fact that a student cannot communicate clearly the code they are suspected of having not written may not mean anything when students under no suspicion also cannot clearly communicate! We must be careful to not prejudge the result, and equally careful not to stack the deck against students.

In the event that a student's work fails all possible good faith considerations, my own preference during hearings of academic misconduct is that the student be provided access to the full annotated transcripts of their code. While to a certain extent this allows an opportunity for students to shape the narrative of their explanation, in most cases the evidence is reasonably cut and dried. All that providing the code ahead of time does in that respect is allow for students to consider the evidence outside of the fraught, and often stressful, environment of an academic misconduct hearing. My own feelings on this matter is that it is much better to hear a considered explanation, even when it may be manufactured. The alternative is an explanation that is a result of stress, worry and the discomfiture that comes from misconduct being alleged. In the latter cases, we cannot reasonably expect that students can acquit themselves under such conditions even in those situations where they may have a reasonable explanation. In none of the academic misconduct hearings I have been responsible for initiating has there been an explanation that made me feel as if the student had been incorrectly targeted. However, if there was such a plausible explanation, I would like to hear it rationally put forward without the additional stresses implied by a formal academic hearing. In some cases, being presented with the evidence alone may be sufficient to make a student acknowledge the work that they submitted was substantively influenced by external sources.

6. CONCLUSION

The lack of any realistically effective mechanism for algorithmically detecting code plagiarism means that even now the process of identifying academic misconduct is one tied up in issues of academic judgment. However, in identifying submissions that have the hallmarks of external influence, we must be careful not to allow our own plagiarism antennae to override our ethical duty of care to our students.

Within software engineering as a discipline, and particularly within the topic of programming, many of the normal conventions of plagiarism simply do not hold – we work within very limited vocabularies and even within limited structural flexibility. The good practice of software engineering too is in many ways an exhortation towards plagiarism – we endorse, as a field, principles of re-use and the application of generalized solutions to problems rather than encouraging individuals to solve them anew each time. Our reference to algorithms, standardized data types, and design patterns creates a powerful impression that we have a preference, as a field, for standard solutions. Our own conventions regarding attribution too are loose and ill-defined, and may even be impossible to honour within complex environments where authorship may be an emergent property. None of this excuses academic misconduct, but it does help situate it within a context that makes it easier to explain.

When we are suspicious of a submission, the process through which we go is often bespoke and ad hoc – we perceive disharmony in submissions, or find things written in ways that are entirely alien to the structure we have inculcated into our students. It is those submissions which most trigger those reactions that are likely to receive the most attention in terms of further investigation, and this has a risk of skewing the results towards those who are least likely to be intentionally attempting to mislead.

We must be mindful then to ensure that the plagiarism investigations that we perform are not only focused where we have suspicions, but also where we have no reason to assume dishonesty at all. Plagiarism which is the best well-executed is almost by definition the least likely to trigger our initial suspicions. As a result we should be careful not to give the clever plagiarists a free ride while we focus our attention on those who simply did not understand the obligations of attribution.

7. REFERENCES

[1] Aiken, A. (2005). Moss: A system for detecting software plagiarism. University of California–Berkeley. [Available from See www.cs.berkeley.edu/aiken/moss. html].

[2] Bartlett, T. (2009). Cheating goes global as essay mills multiply. The Chronicle of Higher Education, 55(28), A1.

[3] Batane, T. (2010). Turning to Turnitin to Fight Plagiarism among University Students. Educational Technology & Society, 13(2), 1-12.

[4] Gullifer, J. M., & Tyson, G. A. (2014). Who has read the policy on plagiarism? Unpacking students' understanding of plagiarism. Studies in Higher Education,39(7), 1202-1218.

[5] Kim, D., Han, Y., Cho, S. J., Yoo, H., Woo, J., Nah, Y., ... & Chung, L. (2013, March). Measuring similarity of windows applications using static and dynamic birthmarks. In Proceedings of the 28th Annual ACM Symposium on Applied Computing (pp. 1628-1633). ACM.

[6] Mahmood, Z. (2009). Contract cheating: a new phenomenon in cyber-plagiarism. Communications of the IBIMA, 10(12), 93-97.

[7] Marsh, B. (2004). Turnitin.com and the scriptural enterprise of plagiarism detection. Computers and Composition, 21(4), 427-438.

[8] Savage, S. (2004). Staff and student responses to a trial of Turnitin plagiarism detection software. In Proceedings of the Australian Universities Quality Forum.

[9] Turner, J. (2012). Developer Week in Review: Are APIs Intellectual Property? In Radar. [Available online at http://radar.oreilly.com/2012/05/api-oracle-googlecopyright-c.html]

[10] Zoubi, Q., Alsmadi, I., & Abul-Huda, B. (2012, May). Study the impact of improving source code on software metrics. In Computer, Information and Telecommunication Systems (CITS), 2012 International Conference on (pp. 1-5). IEEE.

Teaching smart phone ethics: an interdisciplinary approach

Simon Jones
Department of Computer Science
Middlesex University
The Burroughs, London NW4 4BT
Tel: 44-0208-411-4299
s.jones@mdx.ac.uk

ABSTRACT
The phenomenal rise of the smartphone, and the rapid diffusion of mobile computing generally, are amongst the most notable developments of recent times in information and communication technologies (ICTs). The smartphone has become a ubiquitous communication tool, evolving into a digital Swiss Army knife, with an ever growing number of functions, from personal communications manager, navigation system, gaming terminal and camera, to payment device, internet access point and all-round digital lifestyle hub. For these reasons, the smartphone represents a prime topic for teaching and thinking about ICT ethics. This paper proposes an inter-disciplinary approach to this task.

Categories and Subject Descriptors
K.3.2 [**Computers and Education**]: Computer and Information Science Education - *computer science education, curriculum, information systems education*

K.4.1 [**Computers and Society**]*:* Public Policy Issues – *ethics, privacy, regulation, use/abuse of power*

K.5.2 [**Legal Aspects of Computing**]: Governmental Issues - *regulation*

K.7.4 [**The Computing Profession**]: Professional Ethics – *codes of ethics, ethical dilemmas*

General Terms
Design, Economics, Security, Human Factors, Legal Aspects.

Keywords
Smartphone, ICT Ethics, Pedagogy, Inter-disciplinary, Framework

1. INTRODUCTION
This paper draws on several years experience of teaching computer ethics to a culturally heterogeneous body of undergraduate computing students across different global campuses. The approach proposed here emerges partly out of a growing dissatisfaction with the standard approaches to computer ethics pedagogy, articulated in the existing body of textbooks in the field. I have discussed some of the limitations of these approaches in greater detail elsewhere [19]. In brief, multi-faceted technologies like the smartphone threaten to overrun the traditional topic boundaries and theories that underpin many of these texts which draw on quite specific strands of European classical moral philosophy. There is a tendency to present these theories in terms of various oppositions and dichotomies, such as deontological versus teleological, or moral intentions versus consequences. Ethical theories become abstract schema of rules that are applied to specific dilemmas in the ICT field. While ICTs are acknowledged as having a social impact, the complexity of the inter-relationship between technology and society is often lacking. When applied to current, real-world cases in a classroom context, the explanatory power of these classical ethical theories can be limited. They can result in prescriptive approaches that are disembodied from complex scenarios which generate a range of social and ethical issues around ICT. Most problematically, they don't offer much help in resolving these issues or generating feasible practical solutions.

The approach outlined here proposes a revised pedagogic and analytical approach. Rather than laying out the available ethical theories first, and treating the social effects of ICT as an addendum, it places the social and economic context of ICT upfront, methodologically. It then proceeds to explore ethical and legal issues, before concluding with questions of professional practice. In doing so, this approach draws on various theories, including elements of science and technology studies, information systems research, sociology, critical theory and communication and cultural studies. These theories are brought to bear on different moments of the framework to illuminate the different issues generated by a multifaceted phenomenon such as the smartphone.

2. PUTTING SMARTPHONES IN THEIR PLACE: THE SOCIO-ECONOMIC CONTEXT
ICTs don't just appear, or fall out the sky, to land on our desks, or in a shop. It's an obvious, yet important point, in pedagogical terms. ICTs emerge out of particular social and historical contexts. What they look like, how they work, and what they are

used for, are inextricably bound up with those contexts. This much we know from the long traditions of research into technology and society in the disciplines cited above, and their numerous sub-fields [2][23][18][28][38]. From this diverse body of work, we can confidently state that ICTs are always developed and implemented for a purpose, according to particular agendas. They are shaped by fundamental forces and recurring drivers, be they industrial, military or political. Who is funding the research and development of ICTs are powerful influences on the direction of their development, and on their properties and capabilities.

ICTs are always accompanied by social practices and values. These enter into all stages of the software development process and get baked into computer systems [15][18][39]. They are embedded in fine-grained code, algorithms, rules and patterns of reasoning [10]. Engineers make decisions about the architecture of systems and the physical characteristics of devices. Design embodies fundamental assumptions about users, their cognitive abilities and bodies, their imagined needs and wishes. The defaults and options embedded in architectures structure and shape users' choices [40]. In all of these ways, ICTs play a configuring role, shaping the possibilities of what can be done with them by enabling certain options or closing them down, by allowing certain uses while preventing or limiting others. All of these dimensions and properties therefore have an inherently ethical dimension, and ethical analysis requires these embedded values to be disclosed and critically examined [3].

2.1 Smartphone drivers and properties

The smartphone is the product of myriad drivers and shaping forces which have resulted in particular technical properties, discourses and social uses. It is these unique properties, their design, their implementation and their use in specific domains that lie at the heart of many ethical dilemmas raised by smartphones. A grasp of this "big picture" context is therefore a prerequisite to the ethical evaluation of smartphone technology.

This involves some understanding of the mobile phone industry itself, its particular business structures and shaping forces. The precise mix of these elements differs regionally and nationally, according to numerous factors, such as the existing infrastructure for fixed lines, the marketisation of licenses to commercial telecommunications operators, the apportioning of the wireless spectrum, and the role of government policy and regulation [14].

The two major players in the industry, besides the regulating bodies, are the handset manufacturers and the network operators, each with distinct corporate interests and business models. For example, Apple's premier profit engine as a handset manufacturer is the iPhone. With profit margins of 40%-50% per phone, profitability is the primary driver, and Apple's business is built around iPhone sales to network operators and users. Samsung, by contrast, prioritises sales volumes of different models at lower margins [43]. Google's mobile business strategy is built around advertising revenue. Its operating system, *Android*, and the various services and applications that are pre-installed with it are a lucrative advertising space, and a way of targeting and profiling users demographically. As such, there is a strong incentive to collect as much data as possible about users, and Google collects and mines this data in order to improve the accuracy and effectiveness of its advertising services. A common business strategy of all the major smartphone manufacturers is to lock users into proprietary ecosystems of integrated products and services, that include devices, platforms and native apps distributed through online stores.

The network operators' business model is service-based offering contracts to phone users at the retail end, while reselling services to other virtual network operators at the wholesale end. Of these services, prepaid contracts to subscribers are the most profitable, with the trend towards tiered-pricing based on bandwidth consumption, data and bundled services. Network operators also aggregate and sell phone usage data to marketers, advertisers and retailers.

The monetization of the smartphone, and the services and data that flow through it, shape the direction of its development in fundamentals ways. One area where this can be seen is the steady turnover or "churn" of new products, evidenced by the typical lifetime of a phone (18-20 months) and the approximately 1712 phones that are replaced every hour in the UK alone [7]. This is manifested in a continuous drive to create and sell new models with new features, and to push consumers into more expensive and lucrative contracts. This has direct implications for the ways in which smartphones are marketed, and feeds back into the way they are designed and manufactured to incorporate varying degrees of planned obsolescence [36].

Smartphones, however, are not only the result of intellectual work in product design and engineering. Value is also added through physical and mental labour embodied in the construction of the device itself. This can be seen by looking "upstream" in the smartphone's supply-chain, to the production process, and further back to the sourcing of raw materials. Mobile phone components use various mineral elements, chemicals and materials. LCDs, for example, use indium and tin oxides (ITOs) which are by-products of lead and zinc. ITOs are ubiquitous in touch screen devices because of their unique properties. Tantalum, tungsten, tin and gold (3TG) are also critical to the manufacture of smartphones, as is lithium which is used in batteries. ITOs and lithium are rare and difficult to extract, and their production is limited to certain parts of the globe. Mining frequently occurs in politically unstable and/or impoverished countries, such as eastern Congo where the extraction and trade of 3TGs have been controlled by armed militias. The manufacture and assembly of smartphones also occurs in particular parts of the global economy, especially China, where phones are constructed by contract manufacturers at low-cost, high efficiency and high volume, using just-in-time production models. Churn and the frequent launches of new smartphone models invariably mean short delivery times imposed on manufacturers, which in turn have implications for work conditions in these production sites.

Smartphone functionality is dependent on a plethora of interconnected technologies, including the wireless telephone infrastructure of towers, switches, exchanges and cellular grids, as well as wireless protocols and standards. The development of the smartphone is itself predicated on innovations in batteries, miniaturisation and data processing. These have occurred in close parallel with innovations in the wider infrastructure. The addition of a separate Subscriber Identification Module (SIM), for example, is just one example of an innovation that allowed subscription contracts with operators to be separated from the handset device itself. It is the affordability and flexibility of these payment systems which partly explain the rapid diffusion of wireless telephony as a technological and economic substitute for fixed lines, especially in developing countries.

The digitalisation of the wireless infrastructure, and the configuration of the telecommunications network as a whole, have had significant implications for the processing and tracking of data flowing through these networks. Whenever a phone is

powered on and registered with a network, it can be located using triangulation, by analysing the signal strength that different towers observe from that phone. This also gives network operators, and other interested parties, the ability to intercept and record data about calls, devices, SIM cards, and their numerous attributes.

This location awareness and tracking ability was further enhanced by the equipping of smartphones with GPS receivers, by which phones could calculate their position in relation to signals transmitted by satellites. This location data can be transmitted over wireless networks to location-based services, but also to other GPS-receiving devices in the phone's vicinity. Smartphones also have other shorter-range wireless radio transmitters in the form of Wi-Fi and Bluetooth, both of whose signals include a unique, device-specific serial number or address assigned by the manufacturer.

These, and other signals emitted by smartphones, mean that the smartphone is continuously receiving and narrowcasting information about its location and movements. This data can be intercepted and observed by different receivers, then aggregated and analysed to build intelligence about particular phones and their users. Location analysis companies, for example, use strategically positioned devices to locate and track smartphones in retail environments, in order to understand customer behaviour, and send location-based ads to those phones [29].

These capabilities, combined with developments in context-awareness and machine-sensors, have made the smartphone a key point of convergence of wireless and geo-spatial technologies. They have put the smartphone at the centre of emerging networks of smart objects and sensors which are perpetually Internet-connected and communicate wirelessly. As these networks proliferate, mobile computing has extended into a wider range of public, private and domestic environments, endowing physical spaces with the interactive character of the Internet.

2.2 Smartphone language

If ICTs are always accompanied by social values, they are also always accompanied by *discourses* that frame the way they are represented and thought about. Putting ICTs in their social and economic contexts therefore also involves thinking about the *language* of ICT. Critical theory, social scientific and humanities-based approaches can shed light on how these discourses work, through various rhetorical devices, to present a particular set of narratives about technology. They can tell us how these discourses are reproduced, how certain representations of ICT become naturalised, and how these, in turn, serve to maintain particular vested interests and power relations [38]. This kind of critical unpacking and deconstructing of these discourses is an important part of computer ethics pedagogy.

In terms of smartphones, this means looking at the language and imagery used in corporate websites, advertising, and both new and old media. It entails looking at the cultivation of brand worship and the construction of the smartphone as a centrepiece of a consumer culture. These are part of wider discourses of consumerism and "upgrade culture" which pervade the marketing of electronic devices in general, and are a direct consequence of churn. They are rooted in more general technologist narratives about innovation as a process of continuous, linear progress, and the fetishising of the "new". Other recurring discourses that have been identified are ideologies of "speed", "convenience", the need to be perpetually contactable, and the valorising of aesthetic features as a means of expressing individual identity [27]. It is worth trying to foster an awareness of such discourses, not only because they shape common-sense attitudes and "school" us to consume smartphones in certain ways [41], but they also because they feed back into design, development and research.

2.3 Smartphone uses

While these discourses undoubtedly shape the ways in which smartphones are experienced and used, they are also interwoven with a whole range of creative uses and meanings. This much is evident from the swathe of studies in media and communication, and social science, which show how ICTs in general and mobile phones in particular are creatively appropriated in different contexts [14][21][22][37]. These studies show that consumer technologies are always subject to a process of meaning-making by their end users [5]. The used of SMS-based texting is just one example of how phone users have adapted features and developed uses that are not necessarily in the cards of product designers and business strategists. Texting was taken up en masse as a cheaper, alternative mode of communication to voice calls, spurred by the need to optimise messages and reduce the cost of transmission. SMS subsequently evolved into a non-standard "writing orality" with its own vocabularies in different languages [4]. The camera is another example of how smartphone functionalities have been creatively appropriated and incorporated into everyday life. Camera functions, in combination with social networking platforms, have placed the means of image-making and sharing in the hands of smartphone users with various cultural and political implications.

These are just two examples of how users customize smartphones for their own purposes and find innovative uses and workarounds that are often unforeseen by their designers and manufacturers. They suggest that the ways in which smartphone technology is used, by whom, and in what context, is always culturally specific and socially differentiated, and has implications for relations of power, whether in the family, work, or education. The most evident example of this is the central position that the smartphone has come to occupy in youth cultures, globally, where it has become a key tool in the construction of young people's identity, enabling new modes of networked sociability [16]. This is part of a wider process in which mobile telephony has reconfigured communication practices in general by enabling existing networks of relationships and affiliations to be reinforced [4]. This has had positive public safety implications for groups such as the young, the elderly, and the vulnerable, providing an immediate safety link to a personal support infrastructure and to assistance for those in harm's way. Mobile telephony has also enabled new kind of networks and communication flows outside of mainstream media, facilitating the formation of fluid, spontaneous "communities of practice" amongst ad-hoc groups, from flash mobs to political protests [4].

3. ETHICAL PRINCIPLES AND ISSUES

One of the main difficulties that students of ICT ethics have is identifying ethical issues, and explaining *why* these are issues. Ethical issues, to my mind, occur where certain core ethical principles, values or rights are at stake. These issues arise from the particular properties and capabilities of ICTs, and from their design, production, implementation and usage in particular domains. From a teaching perspective, this means teasing out these underlying principles or rights. It means naming them and sourcing them. The ethical issues raised by smartphones touch on a number of core principles and values. As for the sources, these range in scope from broadly-shared human moral values, through

internationally–recognised declarations, treaties and constitutions, to political and moral philosophy, including, but not limited to, the European classical canon.

A useful departure point from which to explore these issues is to reflect on some of the more identifiable controversies related to smartphone *use* as a communication device. While driven partly by media discourses and moral panics about the negative social impact of smartphones, there are issues worth exploring around the consequences of the smartphone's incursion into all areas of public and private life. The familiar scenario of the mobile phone ringing randomly in any given situation, and its potential to disturb or disrupt solitude or concentration—these have highlighted the boundaries of socially acceptable use in different public and private spaces, and touch on wider questions of social etiquette and civility. A related, and oft-noted issue is the phenomenon of "absent presence" where phone users are physically and socially present in any given space, while their attention and mental focus is elsewhere. It is a phenomenon most of us who teach in higher education are probably well familiar with. This touches on a wider problem—the possibility of a communications culture of ***permanent distraction*** being created, one that is decreasing the time available for people to think uninterrupted, at work, at home or in college. Some have suggested that we are becoming so enmeshed in our digital connections that we are neglecting others in our immediate social environment [41].

Another aspect of this redefinition of the boundaries between public and private space is the impact of mobile communication on the work-life balance. Here, the smartphone has become something of a Trojan horse through which work has infiltrated the home. Its "always on" capabilities have helped foster a 24/7 work culture of ***permanent availability*** which threatens the work-home balance in potentially harmful ways. While some of these debates are premised on conjecture and anecdote, the evidence is starting to come in from research in psychology and medicine that heavy smartphone use can detract from inter-personal relationships, interfere with sleep patterns, and lead to higher stress levels [34][35].

Given the large amount of personal data that is narrowcast every time it is switched on, and the nefarious ways in which this data is processed and used, the smartphone has inevitably become a major focus of ***privacy*** concerns. Smartphone capabilities have enabled new kinds of lateral surveillance and privacy incursions *between* citizens, but it is the unprecedented degree of access to the flow of personal information by private and state organisations that is of particular concern. Governments can, and have, forced network operators to turn over location data about users in real-time or as historical records. Concerns have been raised about personal data being gathered in ways that are subject to negligible regulation or oversight. A number of covert surveillance systems, operated by various governments, have been shown to exist, including systems operated by NSA in the United States [PRISM] and GCHQ in the UK [TEMPORA]. These have enabled security agencies to tap into the wireless network infrastructure, and collect metadata, in bulk, about mobile phone use globally. Various techniques for analysing mobile phone usage and call data have been incorporated into these systems. These data analysis tools can be used to determine not only a user's location, but also their historical activities, participation in events, personal beliefs and relationships.

Private corporations also have a major commercial stake in accessing and mining this data. Cellular tower connections, when combined with GPS, wi-fi and other signals represent a powerful dataset that can be used for behavioural profiling and targeted advertising. Passive location services that operate without any clear indication or visibility to users have been particularly contentious [6]. Where users' personal data is gathered, processed and shared between organisations without their knowledge or consent, these privacy questions are closely intertwined with ***data protection*** issues. These scenarios highlight the fundamentally asymmetrical distribution of privacy rights around smartphones. In order to use applications and access services, phone users must enter privately-owned networks which require them to surrender their personal data and consent to varying degrees of monitoring. While users are increasingly transparent to such monitoring, the organisations doing the monitoring are increasingly opaque and protected by a shield of privacy [1].

Smartphones have specific technical vulnerabilities which throw up a number of ***security*** issues. The very nature of wireless radio signals, and their technical properties, makes smartphone communication data vulnerable to interception. Default levels of encryption of transmitted data are relatively weak in both smartphones devices and in the mobile communications network as a whole. Smartphones themselves are particularly susceptible to malware distributed via insecure applications or software updates. Unauthorised access through such malware can be used to read private data, make a phone pretend to power off while remaining on, or activate its sensors and functions (such as the microphone, camera or GPS) in order to monitor the phone's location or immediate environment.

As with many technical threats in the computing field, the ethical issues revolve principally around the *response* to those threats, the adequacy of such responses, and underlying issues raised around responsibility and trust. While security is a key ethical principle and a fundamental right, it is also itself a contested discourse. Tensions exist between users' wishes and demands for appropriate protection and security measures, on the one hand, and corporate priorities around cost on the other. Security is also a commodity that can be exploited economically, invoked to protect certain interests, or used to serve particular agendas and override other legitimate rights, such as privacy and anonymity [39].

Moor's notion of the "invisibility factor" inherent in computer technologies remains as pertinent as ever when thinking about smartphone ethics [24]. The fact that smartphone operations are, to most users, hidden from view, raises some important issues around ***transparency***. Entranced as we are by the seductive, tactile interface of the smartphone, most of us do not fully know how all of its applications and location-based features work. Smartphone technologies, like many ICTs, are "blackboxed," their inner workings opaque to non-technical users. They announce their whereabouts, and they collect and process data, in ways that are invisible to their users. As smartphones become increasingly intelligent, working autonomously in the background, predicting and making decisions on the user's behalf, this is likely to become even more the case.

Many of the systems that run on smartphones are "closed", not reprogrammable and updated remotely by the manufacturers themselves. Access to the underlying code, even in apparently "open source" programs, is partially restricted. 3rd party apps which are developed for Android or iOS are carefully vetted and screened, and can often only be distributed from a manufacturer-maintained online store. Most smartphone devices are

deliberately designed to prevent access to their inner physical workings through the gluing together or encasing of key internal components. This makes them difficult to disassemble and repair.

Some have argued that these features result in "tethered", appliance-like devices which can only be modified on the manufacturer's terms, curtailing the ability to customize, and thereby suppressing innovation and generativity [42]. Compared to desktops and laptops, smartphones give the user much less ***control*** and ***autonomy***. The net result is a device where it is more difficult to replace the operating system, harder to investigate malware attacks, harder to remove or replace undesirable bundled software, more difficult to prevent 3rd parties from monitoring how the device is used and harder to block ads embedded in mobile apps through anti-advertising technology [6]. These issues touch on many of the core principles of the Free and Open Source Software (FOSS) movement and cross over into issues of intellectual property rights.

The status of the "user" in phone design, in the business strategies of network operators, and in regulatory frameworks, is another contested area. Key issues here are the extent to which users are involved in design decisions by manufacturers, or consulted in decisions about policy and regulation. Design assumptions are often based on anecdotal evidence rather than structured engagement with intended users [43]. Here too, major tensions exist between the agendas of phone manufacturers and network operators, on the one hand, and users, on the other, struggling for fairer and cheaper charges, more control over their data, enhanced security, and clearly understandable privacy policies and permission requests. Users struggle against being locked into misleading service contracts in which subscribers are routinely overcharged, resulting in unused capacity for calls and data, and thus surplus profits to network operators. These struggles are manifested, for example, in online campaigns by users to get manufacturers to install "kill switches" on devices to enable data to be erased remotely from stolen phones [6]. They can be seen in struggles around the right to unlock phones from being tethered to a single network, or to "jailbreak" them by obtaining access to their underlying programs and file structures.

The smartphone raises a whole gamut of issues around ***equality***, ***fairness*** and ***inclusion*** at each point in its lifecycle. The rapid diffusion of mobile telephony in developing countries has undoubtedly democratised communication due partly to the proliferation of used and affordable phones, and the lower infrastructural costs of maintaining a cellular tower to serve a whole area compared to laying landline cables into individual households [22]. Examples abound of mobile telephony being used to disseminate public health information, provide access to education, financial services and market information for small businesses [43]. However, it remains unclear to what extent these processes have narrowed the digital divide, or mitigated the disparities in Internet connectivity and access to digital resources, globally.

In those countries with relatively high smartphone adoption rates, it is also unclear what benefits they have brought to those users historically excluded from ICTs, or whether they have simply resulted in new forms of exclusion. With smaller screens and keyboards, and slower connections compared to desktop-based, wired, broadband computing, some have argued that smartphones represent a cheaper, 2nd tier of access. Smartphone-based paradigms of computing are less conducive to creating content, and unsuited to many forms of computer-based productive work [43]. There are questions marks too around the extent to which smartphones have benefitted the elderly, or groups with impaired cognitive, sensory and physical abilities. This raises design issues around the usability of touch-screen interfaces, and the navigability and accessibility of displays and input functions.

Equality issues also arise at both ends of the smartphone's supply chain around the human cost of raw materials extraction, manufacturing and recycling. Where these processes are carried out under hazardous, exploitative or inhumane conditions, or where they serve to exacerbate conflict and suffering, there are serious humanitarian issues involved.

Finally, there are ***environmental*** issues at each point in the smartphone's lifecycle. Many of the chemicals, elements and materials contained in smartphones and their components are either finite, toxic, carcinogenic, or all three. Where the extraction of such materials results in mineral depletion, toxic waste or large spoil heaps, there are issues of sustainability and environmental harm [25][26]. In terms of the smartphone's carbon footprint, most of its energy consumption and CO_2 emissions occur in its manufacturing and usage. Mobile-to-mobile calls use three times more power than landline-to-landline calls [7]. For a single smartphone, the energy used to transmit calls across a wireless network over a 1 year period, is equivalent to three times the CO_2 emissions involved in its manufacture [43]. At the disposal end of the lifecycle, unregulated recycling also poses hazards to both workers and to the environment through the handling of toxic waste, and its accumulation in dumps and landfills.

4. SMARTPHONE LAWS AND REGULATIONS

The law is an important touchstone for both prospective and existing IT professionals. Knowledge of the relevant legislation in any given issue is a crucial part of computer ethics, as is legal compliance in the evaluation of solutions to particular dilemmas. Like areas of new and emerging technologies, however, there is a relative lack of legal and regulatory frameworks governing smartphones per se. The law, with its comparatively gradual pace of legislative debate and enactment, is generally behind the curve of innovation in smartphone technology.

Most countries have government bodies that regulate the telecommunications sector, for example the FCC in the USA, and OFCOM in the UK. In the UK, there is statutory legislation that prohibits the use of hand-held mobile devices while driving in the form of a 2003 amendment to the *The Road Vehicles (Construction and Use) Regulations.* The existing legislation that pertains to smartphones is focussed around data protection, intellectual property, electronic waste, and the sourcing of conflicting materials. Regarding intellectual property, there have been significant legal disputes about corporate control of patented elements of smartphone technology, and the rights to exploit these, most notably between Apple and Samsung. The collection, treatment and recycling of phones is regulated by the EU's *Waste Electrical and Electronic Equipment* (WEE) directive, 2002/2012. The USA's *Dodd–Frank Wall Street Reform and Consumer Protection Act*, 2010 obliges companies to disclose conflict minerals from the eastern Congo in their supply chains, and to remove illegally mined minerals from them. The EU's *Privacy and Electronic Communications* of 2002 extended the EU's *Data Protection Directive* of 1995 to include prohibition of unsolicited texts and messages distributed to mobile phones. In the area of privacy, UK government proposals under the *Investigatory*

Powers Bill 2015 would require mobile operators to log their customers' call data, and provide government access to that data.

While legal compliance is an important benchmark of professional practice, on its own, it is an insufficient guarantor of ethical design, implementation or use of smartphone technologies. The law has a number of limitations, around issues of jurisdiction, enforcement and effectiveness that need to be explored. The applicability of EU data protection legislation to US-owned global corporations doing business in Europe remains an ongoing point of legal contention, with Google and others lobbying for EU privacy laws to be relaxed. Existing data protection principles enshrined in the 1995 EU data protection framework are put to the test by smartphone data, particularly around informed consent, disclosure to 3rd parties and data retention. Much of the data that flows through, and is stored on, smartphones, and associated cloud services, could rightly be considered "sensitive" given that it represents user's thoughts, habits, locations and movements. Laws are not necessarily ethical, nor are they politically or economically neutral. Some laws are weighted in favour of users' rights, while others tend to protect the vested interests of private corporations or those of the state. Laws can also be circumvented and loopholes exploited, be they regulations on recycling or hardware disposal, or reporting on environmental impacts. Phone manufacturers and networks, for example, attempted to delay and weaken *The Dodd Frank Ac*t through their corporate lobbyists and trade associations. Laws and regulatory frameworks therefore need to be critically scrutinized, and the issues that they raise explored. Some ethical issues, it needs to be acknowledged, cannot and perhaps should not, be solved necessarily by statutory or regulatory interventions.

5. DOING THE RIGHT THING: SMARTPHONES AND PROFESSIONAL PRACTICE

The approach outlined in this paper is grounded in an *applied* definition of ethics, one which considers the ethical issues raised by ICTs with a view to informing practice and illuminating potential solutions to those issues. A key aim of this task is therefore to look at the implications of the preceding three stages for professional practice. The end goal of the analysis, in this sense, is the practitioner moment. First and foremost, this involves looking at the codes of conduct of relevant professional bodies, and to what extent their standards of practice are applicable to practitioners in the smartphone domain. How can professional responsibilities be balanced with the rights of different stakeholders, with budgetary and time constraints, considerations of technical feasibility, functionality and aesthetics, and with all the drivers and forces which impinge on individual practitioners? This difficult balancing act needs to be explored, while simultaneously acknowledging some of the limitations of professional codes of conduct in resolving the social and ethical issues raised above.

It is useful, at this point, to widen the notion of ethical responsibility beyond questions to do with *individual* professionals, to those which have implications for *organisations*, be they private corporations or government agencies. This means scrutinising the codes of practice and mission statements of companies operating in the smartphone industry. To what extent do their actions and deeds measure up to their public statements and policies, particularly in areas such as environmental impact, privacy and transparency? To what extent are organisations transparent about their operations, whether government agencies about their monitoring and surveillance practices, or phone manufacturers about their supply chains and their environmental impacts? Audits of the latter reveal that most are not living up to their claims, while disclosures about the former reveal a major lack of transparency and independent governance [25][26][7]. Where public pronouncements about ethical goals are not fulfilled or contradicted by factual evidence, companies run the risk of courting unwelcome public scrutiny, boycotts and legal action, resulting in reputational damage and potential loss of business.

Important as it is to identify cases where ethical principles are threatened, whether by unethical design, production or use of ICTs, ethical analysis also needs to provide a vision of what "good" looks like in practical terms. It is important, in this sense, to propose solutions and alternatives, and to imagine how things might be different. How can smartphones be designed in ways that *affirm* principles of privacy, autonomy, transparency and inclusion? How might these principles be embedded in the development process and translated into procedures that can be followed by programmers and engineers in real-life projects? Answering these questions is beyond the scope of this paper, but I'd like to conclude by offering some pointers and concrete examples of how these principles should, and indeed already have, been put into practice.

User-centricity has been repeatedly affirmed as a key principle that should inform the entire ICT development lifecycle, from requirements gathering to evaluation and testing. Value-sensitive design entails the involvement of key stakeholders and prospective users in the design process from the outset [9]. These approaches provide a way of incorporating principles of autonomy and transparency into each stage of the development lifecycle. ICT development, in this sense, should not just be the result of technology "push," but also participation and involvement of users and the broader communities of which they are a part [30].

Principles of sustainability and environmental protection should be implemented throughout the smartphone lifecycle, commencing with the use of alternative raw materials in product design and manufacturing. This also implies the sustainable use and recycling of *existing* materials in order to mitigate the depletion of non-renewable resources. It means green procurement of components which don't use toxic chemicals and materials, and which in turn don't require extraction of rare earths which involve toxic waste or the use of conflict materials. Principles of sustainability might also entail using alternative, organic or bio-degradable casing materials, exploring alternative sources of battery power, or battery-less phones which derive their power from radio signals or solar energy, or which harvest energy from physical movement in everyday human activities though new types of fabric [12]. Reducing the environmental burden throughout the supply chain also means regulated, transparent and clean disposal and recycling.

Overall, this implies moving away from paradigms of ICT design which are founded on disposability, built-in obsolescence and the upgrade culture of "fast tech" towards new kinds of "slow tech" design which are "clean", "good" "fair" and "open" [30]. "Fair" in terms of ensuring that working conditions throughout the supply chain are humane and non-exploitative; "good" in helping people find an appropriate balance between work time, free time and leisure; "open" through innovation and development founded on openly defined standards and architectures which others can adapt and freely improve upon; "slow" in terms of slowing down the ICT lifecycle and turnover of devices through a greater focus on modular products which enable components, rather than whole

devices, to be replaced, and a greater emphasis on repair and re-use. Such models are also "responsible" not only through greater accountability and transparency in the innovation lifecycle, but also through greater public participation and engagement, and more interaction between innovators and end-users, [39]. The following examples provide some brief glimpses of these principles in practice.

Social enterprise smartphone manufacturer *Fairphone* is founded on transparency about its business operations, and uses supply chains that aim to be free of conflict materials. The production of its first smartphone was financed through online crowd-funding [8]. *Modular* smartphones are designed to be upgradable through the insertion of small plug-and-play modules into a smartphone shell. These enable functionality to be added, removed or adapted according to use or context, such as wi-fi connectivity, large screens, cameras, speakers and processors. Examples of modular phones include *phonebloks* [3] and prototypes developed by Google's Advanced Technology and Projects division [11]. *Privacy-enhancing* features that are built into smartphones can provide different levels of privacy and security for different services, and greater protection against rogue apps. These give users greater control over permission requests at both install and run-time, along with the ability to block access to certain phone functions, location services or personal data. Google's "Apps Ops", for example, was designed to be incorporated into its Android M software and allows users to pick and choose which data and functions apps have access to, on a case-by-case basis [13]. Security smartphones, such as the *Quasar IV* cipherphone use self-authenticated verification, bio-metrics and asymmetric strong encryption to safeguard users' digital identity [32]. Online services such as the wiki-based website *iFixit*, allow users to create, edit and share repair manuals for smartphones. *iFixit* uses teardowns and reverse engineering to openly share technical knowledge amongst smartphone users [17]. Finally, local social enterprises, such as the London-based *Restart Project,* focus on extending the lifespan of smartphones through repair and resilience. *Restart* promotes a waste-nothing "circular economy" and encourages people to use their electronic devices longer, by sharing repair and maintenance skills [33].

6. CONCLUSION

This paper has outlined a revised framework for ICT ethics teaching, and illustrated this framework by applying it to the smartphone. This approach consists of four stages of analysis, each driven by a particular set of key questions, which, when combined, provide a holistic multi-dimensional framework, that can be applied to ICTs across their lifecycle. As mobile computing becomes more ubiquitous, intelligent and embedded in everyday life, so its ethical implications cannot be fully grasped within the confines of any single discipline. Phenomena such as smartphones cross over the standard topics and ethical theories used in many existing computer ethics frameworks. This paper points to the potential value of an inter-disciplinary approach which draws on varied theoretical tools with different explanatory strengths, enabling new connections and insights to be generated across disciplinary boundaries. From a teaching perspective, the framework outlined in this paper provides students with a flexible methodology for doing ethics themselves, and a means to explore the ethical issues raised by *any* ICT, in any domain or topic area of interest. This paper suggests that the evaluation of ethical courses of action and potential solutions can be enriched when founded on a deeper understanding of the social and economic contexts in which ICTs are designed, implemented and used. On this basis, the framework has potential relevance not only to students and teachers of ICT ethics, but also to practitioners. How smartphones develop in the future remains to be seen, but the trajectory of that development is by no means pre-fixed or given. The direction of travel lies partly in the hands of our students as prospective future professionals. This approach is a reminder to them, and to us, that how ICTs are designed, made and used, are fluid and mouldable. They are not set in stone, but subject to change and up for grabs.

7. REFERENCES

[1] Andrejevic, M. 2007. *iSpy: surveillance and power in the interactive era*. University Press of Kansas, Lawrence, Kan.

[2] Bijker, W.E., Hughes,T.P. and Pinch,T.J., Eds. 1987. *The social construction of technological systems: new directions in the sociology and history of technology*. MIT Press, Cambridge, MA.

[3] Brey, P. 2010. Values in technology and disclosive computer ethics. In *Cambridge handbook of information and computer ethics*, L. Floridi, Ed. Cambridge University Press, Cambridge, 41-58.

[4] Castells, M. 2009. *Communication Power*. Oxford University Press, Oxford.

[5] Du Gay, P., Hall, S., Janesd, L., Koed Madsen, A., MacKey, H. and Negus, K. 2013. *Doing Cultural Studies: the story of the Sony Walkman* (2nd edition). Sage, London.

[6] Electronic Frontier Foundation. 2015. The Problem with Mobile Phones. *Electronic Frontier Foundation*. https://ssd.eff.org/en/module/problem-mobile-phones

[7] Ethical Consumer. 2013. Mobile phones and broadband. *Ethical Consumer*. Nov/Dec. www.ethicalconsumer.org.

[8] Fairphone.com. 2015, Fairphone. http://www.fairphone.com

[9] Friedman, B., Kahn, P. and Borning, A. 2008. Value Sensitive Design and Information Systems. In *The Handbook of Information and Computer Ethics*, K. Himma and H. Tavani, Eds. Wiley-Blackwell, Chichester, 69–102.

[10] Fuller, M., Ed. 2008. *Software studies: a lexicon.* MIT Press, Cambridge, MA.

[11] Gibbs, S. 2015.Google to launch modular smartphone with switchable parts. *The Guardian*, 15th January. http://www.theguardian.com/technology/2015/jan/15/google-modular-smartphone-switchable-parts-project-ara

[12] Gibbs, S. 2015. Google Atap: touch-sensitive jeans, tiny radar and the death of the password. *The Guardian*, 1st June. http://www.theguardian.com/technology/2015/jun/01/google-atap-io-touch-sensitive-jeans-tiny-radar

[13] Gibbs, S. 2015. Why it took us so long to match Apple on privacy – a Google exec explains. *The Guardian*, 9th June. http://www.theguardian.com/technology/2015/jun/09/google-privacy-apple-android-lockheimer-security-app-ops

[14] Goggin, G. 2006. *Cell phone culture: mobile technology in everyday life*. Routledge, London.

[15] Gotterbarn, D. 2000. Value free software engineering: a fiction in the making. http://csciwww.etsu.edu/gotterbarn.

[16] Green, N. and Haddon, L. 2009. *Mobile communications: an introduction to new media*. Berg, Oxford.

[17] ifixit.com. 2015. iFixit. http://www.ifixit.com

[18] Johnson, D. 2009. *Computer ethics: analyzing information technology*. Pearson, Upper Saddle River, NJ.

[19] Jones, S. 2015. Doing the right thing: computer ethics pedagogy revisited, *Journal of Information, Communication and Ethics in Society*. To appear.

[20] Jouhans, S. 2015. Energy harvesting could be the future of mobile power. *The Guardian*, 4th June. http://www.theguardian.com/media-network/2015/jun/04/energy-harvesting-future-mobile-charging

[21] Katz, J. 2006. *Magic in the air: mobile communication and the transformation of social life.* Transaction Books, London.

[22] Ling, R. and Donner, J. 2009. *Mobile communication*. Polity Press, Cambridge.

[23] MacKenzie, D. and Wajcman, J. 1999. Introductory essay. In *The social shaping of technology*. D. MacKenzie, and J. Wajcman, Eds. Open University Press, Buckingham, 3-27.

[24] Moor, J. 1985. What is computer ethics? *Metaphilosophy*, 16, 266-275.

[25] Monbiot, G. 2013. Smart Phones, Dumb Companies, *The Guardian*, 13th March. http://www.monbiot.com/2013/03/11/smart-phones-dumb-companies/

[26] Monbiot, G. 2013. Apple Turnover, *The Guardian*, 23rd September. http://www.monbiot.com/2013/09/23/apple-turnover/

[27] Nayar, P.K. 2010. *An introduction to new media and cybercultures*. Wiley-Blackwell, Chichester.

[28] Nissenbaum, H. 1998. Values in the design of computer systems. *Computers and society*, March, 38-39.

[29] Path Intelligence. 2015. What we do. *Path Intelligence*. http://www.pathintelligence.com/what-we-do/decision-science

[30] Patrignani, N. and Whitehouse, D. 2014. Slow tech: a quest for good, clean and fair ICT. *Journal of Information, Communication and Ethics in Society*, 12, 2, 78-92.

[31] Phoneblocks.com. 2015. About phoneblocks. *Phoneblocks*. http://phoneblocks.com/about-phonebloks

[32] Qsalpha.com. 2015. Quasar IV. *QSAlpha*. http://qsalpha.com/en/quasar-iv

[33] Restart, 2015. Let's fix our relationship with electronics. *Restart Project*. http://therestartproject.org

[34] Sample, I. 2014. Are smartphones making our working lives more stressful? *The Guardian*, 18th September http://www.theguardian.com/technology/2014/sep/18/smartphones-making-working-lives-more-stressful

[35] Siddique, H. 2015. Smartphones are addictive and should carry health warning, say academics. *The Guardian*, 4th March.http://www.theguardian.com/technology/2015/mar/04/smartphones-addictive-make-people-narcissistic-say-academics

[36] Slade, G. 2006. *Made to break: technology and obsolescence in America.* Harvard University Press, Cambridge, MA.

[37] Snickars, P. and Vonderau, P. Eds. 2012. *Moving data: the iPhone and the future of media.* Columbia University Press, New York, NY.

[38] Stahl, B.C. 2008. *Information systems: critical perspectives*. Routledge, London & New York.

[39] Stahl, B.C., Eden, G., Jirotka, M. and Coeckelbergh, M. 2014. From computer ethics to responsible research and innovation in ICT: the transition of reference discourses informing ethics-related research in information systems. *Information and Management*, 51, 6, 810–818.

[40] Thaler, R. H., Sunstein, C. R. and Balz, J. P. 2010. Choice Architecture, *Social science research network*, http://ssrn.com/abstract=1583509.

[41] Turkle, S. 2011. *Alone together: why we expect more from technology and less from each other*. Basic Books, New York.

[42] Zittrain, J. 2008. *The future of the Internet: and how to stop it*. Yale University Press, New Haven, CT.

[43] Woyke, E. 2014. *Smartphone: anatomy of an industry*. The New Press, London & New York.