Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Generalization of Classic Question Order Effects across Cultures

Stark , Tobias; Silber, Henning; Krosnick, Jon A.; Blom, Annelies G.; Aoyagi, Midori ; Belchior, Ana ; Bosnjak, Michael ; Clement, Sanne Lund; John, Melvin ; Jónsdóttir, Guðbjörg Andrea ; Lawson, Karen ; Lynn, Peter ; Martinsson, Johan ; Shamshiri-Petersen, Ditte; Tvinnereim, Endre ; Yu, Ruoh-rong

Published in:
Sociological Methods & Research

DOI (link to publication from Publisher):
10.1177/0049124117747304

Creative Commons License
CC BY-NC 4.0

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Stark , T., Silber, H., Krosnick, J. A., Blom, A. G., Aoyagi, M., Belchior, A., Bosnjak, M., Clement, S. L., John, M., Jónsdóttir, G. A., Lawson, K., Lynn, P., Martinsson, J., Shamshiri-Petersen, D., Tvinnereim, E., & Yu, R. (2020). Generalization of Classic Question Order Effects across Cultures. *Sociological Methods & Research*, *49*(3), 567-602. Advance online publication. https://doi.org/10.1177/0049124117747304

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Check for updates

Article

Sociological Methods & Research
2020, Vol. 49(3) 567-602
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0049124117747304
journals.sagepub.com/home/smr
SAGE

Generalization of Classic Question Order Effects Across Cultures

Tobias H. Stark[1], Henning Silber[2], Jon A. Krosnick[3], Annelies G. Blom[4], Midori Aoyagi[5], Ana Belchior[6], Michael Bosnjak[7], Sanne Lund Clement[8], Melvin John[9], Guðbjörg Andrea Jónsdóttir[10], Karen Lawson[11], Peter Lynn[12], Johan Martinsson[13], Ditte Shamshiri-Petersen[8], Endre Tvinnereim[14], and Ruoh-rong Yu[15]

[1] Utrecht University/ICS, Utrecht, the Netherlands
[2] Department of Survey Design and Methodology, GESIS–Leibniz Institute for the Social Sciences, Mannheim, Germany
[3] Departments of Communication, Political Science, and Psychology, Stanford University, Stanford, CA, USA
[4] Department of Political Science and Collaborative Research Center 884 "Political Economy of Reforms," University of Mannheim, Mannheim, Germany
[5] Center for Social & Environmental Systems Research, National Institute for Environmental Studies, Japan
[6] Department of Political Science and Public Policies, Instituto Universitário de Lisboa (University Institute of Lisbon), ISCTE-IUL, Lisbon, Portugal
[7] ZPID - Leibniz Institute for Psychology Information, Trier, Germany
[8] Department of Political Science, Aalborg University, Aalborg, Denmark
[9] Collaborative Research Center 884 "Political Economy of Reforms," University of Mannheim, Mannheim, Germany
[10] Social Science Research Institute, University of Iceland, Reykjavík, Iceland
[11] Department of Psychology, University of Saskatchewan, Saskatoon, Canada
[12] Institute for Social and Economic Research, University of Essex, Colchester, UK
[13] Department of Political Science, University of Gothenburg, Gothenburg, Sweden
[14] Uni Research Rokkan Centre, Bergen, Norway
[15] Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

Corresponding Author:
Tobias H. Stark, ERCOMER (European Research Centre on Migration and Ethnic Relations), Utrecht University/ICS, Padualaan 14, 3584 CH Utrecht, the Netherlands.
Email: t.h.stark@uu.nl

Abstract

Questionnaire design is routinely guided by classic experiments on question form, wording, and context conducted decades ago. This article explores whether two question order effects (one due to the norm of evenhandedness and the other due to subtraction or perceptual contrast) appear in surveys of probability samples in the United States and 11 other countries (Canada, Denmark, Germany, Iceland, Japan, the Netherlands, Norway, Portugal, Sweden, Taiwan, and the United Kingdom; $N = 25{,}640$). Advancing theory of question order effects, we propose necessary conditions for each effect to occur, and found that the effects occurred in the nations where these necessary conditions were met. Surprisingly, the abortion question order effect even appeared in some countries in which the necessary condition was not met, suggesting that the question order effect there (and perhaps elsewhere) was not due to subtraction or perceptual contrast. The question order effects were not moderated by education. The strength of the effect due to the norm of evenhandedness was correlated with various cultural characteristics of the nations. Strong support was observed for the form-resistant correlation hypothesis.

Keywords

survey methods, questionnaire design, question order effects, cross-cultural, perceptual contrast

Introduction

Over many decades, much research has yielded recommendations about how to design survey questions optimally (Krosnick and Presser 2010; Saris and Gallhofer 2014; Schuman and Presser 1981; Tourangeau, Rips, and Rasinski 2000). Most of these recommendations are based on the results of experimental studies in which respondents were randomly assigned to be asked a question or set of questions in different ways. Such manipulations have varied the number of points on rating scales, for example, the wordings of questions, the order in which response options are presented, the order in which questions are asked, whether and how to offer a "don't know" option to respondents, and more. The impacts of such manipulations on responses are called "response effects."

One especially interesting type of response effect involves question order. Answers to a target question can be changed by asking another question before asking the target question (e.g., McFarland 1981; Schuman, Presser,

and Ludwig 1981). For example, in one experiment conducted in 1948, 37 percent of Americans said that Communist news reporters should be allowed into the United States to report information to their home countries, whereas 73 percent of other respondents expressed this opinion after first being asked whether a communist country like Russia should admit American reporters (Hyman and Sheatsley 1950). A wide array of different question order effects have been documented in the published literature, and they seem to occur due to a variety of cognitive processes including perceptual contrast, subtraction, and more (Moore 2002; Tourangeau and Rasinski 1988).

However, most of this research on question design has been based on data collected from Americans using questionnaires in English. Only in recent years has an increasing number of experiments been conducted elsewhere in the world. For example, questionnaire design experiments are now routinely administered via the European Social Survey in many countries (e.g., Saris and Gallhofer 2011). But to date, no literature has evaluated the generalizability of well-known question order effects originally documented with U.S. data across a wide array of other nations that differ in terms of language, culture, context, and other attributes. In light of currently escalating desire for evidence on replicability, such research seems particularly valuable (Bollen et al. 2015).

In this article, we report tests of two question order effects that were described by Schuman and Ludwig (1983) and Schuman and Presser (1981): one involving financial contributions by businesses and labor unions and the other involving abortion. We assessed (1) whether the same effects were observed when the same experiments were repeated via Internet data collection 35 years after the original studies in the United States, (2) whether the same effects were observed when the businesses and labor unions experiment was conducted in 9 other countries (11 samples) and the abortion experiment was conducted in 11 other countries (14 samples) after the questions were translated into other languages, and (3) whether the question order effects were stronger among people with less education (Narayan and Krosnick 1996).

The Contributions Experiment

Schuman and Ludwig (1983) identified question order effects that occurred when the second question in a sequence activates the so-called norm of evenhandedness in the mind of the respondent, which changes how he or she answers that second question. In one such study conducted in 1947, U.S. respondents were asked two questions about whether they thought that two different and traditionally antagonistic groups, labor unions and

businesses, should be allowed to make financial contributions to political campaigns (Schuman and Ludwig 1983). When each respondent was asked just one of these questions, people expressed support for contributions by unions more often than those by businesses: 23 percent versus 14 percent, respectively.

When the respondents were next asked about the other group, Schuman and Ludwig (1983) speculated, respondents are likely to have thought about a consideration that would not necessarily enter their minds when asked about only a single group. Specifically, respondents might recognize that unions and businesses are often in conflict with one another, so if one group is given a right or opportunity, fairness would demand that the other group be given the same right or opportunity in order to avoid bias or favoritism. As a result, Schuman and Ludwig (1983) speculated, respondents may shift their answers to the second question in the sequence in the direction of their answers to the question that they had already answered. In line with this reasoning, endorsement of contributions by labor unions decreased from 23 percent when the question was asked first to 16 percent when the question was asked second ($\chi^2 = 17.62$, $p < .001$; see row 1 in Table 1), and endorsement of contributions by businesses increased from 14 percent to 24 percent ($\chi^2 = 47.05$, $p < .001$) when that question was asked second.

This logic suggests that a question order effect inspired by the norm of evenhandedness will only be observed when three conditions are met. First, the norm of evenhandedness must be endorsed in principle by individuals participating in the survey. We would not expect to see such a question order effect among respondents who do not endorse such a norm. Second, respondents must have contrasting evaluations of the two parties mentioned in the question. That is, respondents must normally be inclined to favor one of the groups over the other, so that answers to the questions asked first will reflect that preference, and that preference will be minimized when answering the second questions after activation of the norm. Thus, the question order effect involving the Communist and American reporter questions (Hyman and Sheatsley 1950) probably hinges on the negative evaluations of communist countries and positive evaluations of the United States that were prevalent in America in 1948. Third, respondents must not think of the norm when asked the first question and must be prompted to think of it for the first time when asked the second question. That is, such question order effects might not occur if respondents have enough cognitive capacity to spontaneously think of the norm of evenhandedness when asked the first question.

In line with the third condition, past studies have found that less educated respondents were more likely to manifest question order effects due to the

Table 1. Question Order Effects in the Contributions Experiment in the Original Publication and 11 Samples.

	Businesses Can Make Contributions					Unions Can Make Contributions				
Country	Percentage Yes When Asked First	Percentage Yes When Asked Second	Difference	χ^2	*N*	Percentage Yes When Asked First	Percentage Yes When Asked Second	Difference	χ^2	*N*
Original result										
United States (Gallup) 1947[a]	13.3	23.6	10.3	47.05***	2,634	23.2	16.3	−6.9	17.62***	2,634
Meet necessary condition										
United States (Gallup)	37.0	43.1	5.1	7.85**	1,975	43.4	37.8	−5.6	6.31*	1,974
Canada	17.5	23.4	5.9	7.03**	1,315	24.1	18.7	−5.4	5.64*	1,314
The Netherlands	9.0	13.2	4.2	9.77**	2,234	13.6	14.6	1.0	0.48	2,238
United Kingdom	19.8	25.0	5.2	8.46**	2,159	27.8	25.8	−2.0	1.12	2,147
Necessary condition reversed										
Denmark	46.1	47.8	1.7	0.34	1,305	33.0	38.2	5.2	3.92*	1,307
Taiwan	20.7	19.3	−1.4	0.25	790	9.8	15.0	5.2	4.92*	790
Iceland	8.6	12.9	4.3	14.35***	2,963	4.3	4.2	−0.1	0.03	3,001
Do not meet necessary condition										
Germany (German Internet Panel)	14.2	11.7	−2.5	1.48	1,048	13.1	12.8	−0.3	0.04	1,050
Germany (GESIS)	12.2	11.4	−0.8	0.67	4,206	11.2	12.9	1.7	2.67	4,199
Norway	20.6	23.4	2.8	1.26	1,167	25.0	22.8	−2.2	0.75	1,166
Sweden	20.1	23.0	2.9	2.18	1,718	19.6	21.1	1.5	0.61	1,719

[a]United States Gallup 1947 is the original result presented by Schuman and Ludwig (1983).
*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

norm of evenhandedness (Narayan and Krosnick 1996). Education may moderate the effect because more educated respondents are more likely to think of the norm spontaneously when answering the first question, so norm activation does not occur only when the second question is asked (Schuman and Ludwig 1983). In contrast, less educated individuals may be less likely to spontaneously think of the norm initially and may therefore be more disrupted by its activation when asked the second question.

If this is true, then any attempt to replicate an experiment involving the norm of evenhandedness in the United States might fail if Americans are now more cognitively skilled than they were at the time of Schuman and Presser's (1981) experiments. And indeed, Americans are now much more educated than they were in 1947 (Nie, Golde, and Butler 2009). If this increase in education has been paralleled by an increase in cognitive skills, then perhaps question order effects due to the norm of evenhandedness may now be weaker than they used to be.

In line with this reasoning, Schuman and Presser (1981) found a considerably smaller question order effect in the Communist and American newspaper reporters experiment in 1980 than had been observed in 1948, a time period during which educational attainment in America rose. However, Nie and colleagues (2009) have argued that the apparent rise in education was simply due to a change in the criteria used for credentialing, not a change in Americans' cognitive skills. Therefore, question order effects based on the norm of evenhandedness may be no weaker today than they were decades ago.

The Abortion Experiment

Another well-known question order effect involves abortion. Specifically, Schuman and Presser (1981) discovered that support for abortion being legal when a married woman does not want any more children was 12.6 percentage points lower when that question was asked after respondents were asked about whether abortion should be legal if there is a strong chance of serious defect in the baby ($\chi^2 = 9.52, p < .01$; row 1 in Table 2). This question order effect was not symmetric: asking first about abortion by the women who does not want any more children did not alter expressed attitudes toward abortion in the case of a birth defect ($\Delta = -1.0$ percent, $\chi^2 = .11, p > .05$; Schuman and Presser 1981).

Two different explanations have been proposed for this question order effect. The first involves a change of meaning of the second question. According to Schwarz and Bless' (1992a) inclusion–exclusion model, respondents reinterpret a later question in a sequence in light of inferences about what they think the researcher intends to ask. In this instance, conversational norms are

Table 2. Question Order Effects in the Abortion Experiment in the Original Publication and 14 Samples.

	Abortion When No More Children					Abortion When Genetic Defect				
Country	Percentage Yes When Asked First	Percentage Yes When Asked Second	Difference	χ^2	*N*	Percentage Yes When Asked First	Percentage Yes When Asked Second	Difference	χ^2	*N*
Original result										
United States 1979[a]	60.7	48.1	−12.6	9.52**	293	84.0	83.0	−1.0	0.11	305
Meet necessary condition										
United States (Gallup)	65.0	54.8	−10.2	20.96***	1,963	72.9	74.3	1.4	0.47	1,963
United States (TESS)	56.3	50.8	−5.5	3.06	1,015	69.2	72.7	3.5	1.49	1,015
Canada	80.1	72.1	−8.0	11.48**	1,309	88.0	87.9	−0.1	0.01	1,304
The Netherlands	72.7	60.8	−11.9	35.72***	2,243	91.4	90.2	−1.2	0.95	2,246
United Kingdom	76.8	64.1	−12.7	41.84***	2,183	89.4	89.9	0.5	0.17	2,202
Iceland	86.2	75.8	−10.4	51.99***	2,984	95.1	96.2	1.1	2.13	2,947
Germany (German Internet Panel)	80.2	59.6	−20.6	53.55***	1,048	91.5	91.8	0.3	0.03	1,046
Germany (GESIS)	77.7	56.9	−20.8	205.44***	4,188	89.4	90.1	0.7	0.58	4,172
Portugal	66.4	52.0	−14.4	25.92***	1,204	88.4	90.5	2.1	1.48	1,204
Taiwan	77.3	66.3	−11.0	11.74**	789	94.0	94.6	0.6	0.13	788
Japan	41.2	42.9	1.7	0.43	1,471	67.0	66.7	−0.3	0.01	1,451
Do not meet necessary condition										
Denmark	91.5	81.6	−9.9	27.35***	1,308	91.4	92.1	0.7	0.21	1,302
Norway	85.9	75.2	−10.7	28.66***	1,584	85.9	89.4	3.5	4.30*	1,579
Sweden	93.4	90.0	−3.4	6.64*	1,718	95.3	95.0	−0.3	0.11	1,718

[a]United States 1979 is the original result presented by Schuman and Presser (1981).
*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

thought to dictate that a researcher would not ask the same question twice, just as during an everyday conversation, one participant would not ask, "How are you doing?" immediately after his or her conversational partner has answered exactly that question (Grice 1975; Schwarz 1994).

If respondents are first asked about abortion by a married woman, the respondents might assume that one reason she might not want more children is if the baby might have a birth defect. But if those respondents were first asked about whether abortion should be permitted when a serious birth defect is possible, the respondents may infer that the researcher means the married woman does not want more children for other reasons, not including a serious chance of a birth defect. Assuming that people generally see a high likelihood of a serious birth defect as a compelling reason to permit abortion, then subtracting that reason out from the set of reasons for which the married woman might not want more children may leave fewer respondents supportive of the woman's right to obtain an abortion (Schuman et al. 1981).

Bishop, Oldendick, and Tuchfarber (1985) reported evidence challenging the claim that subtraction is responsible for this question order effect. These researchers asked respondents why they favored or opposed making abortion legal for a married woman who does not want more children. None of the respondents mentioned the possibility of a birth defect as underlying their reasoning.

A second possible explanation for the question order effect is quite different: perceptual contrast. A great deal of research in psychology suggests that perceptions of objects change when perceived in relation to specific other objects (e.g., Higgins and Lurie 1957; Stevens 1957). When lifting a moderately sized, medium-weight block of metal, a person may perceive it to be moderately heavy. But if the person picks up that weight just after previously lifting a very heavy weight, the moderate-sized weight will seem light. By the same token, lifting the moderate-sized weight after lifting a very light weight will make the former seem heavier.

This logic may also apply to Schuman and Presser's (1981) abortion question order effect. If a person perceives a high likelihood of a serious birth defect as a very compelling reason to permit abortion, answering a question about that reason first might make wanting no more children seem to be not as compelling a reason as it would have been if considered alone (Bishop et al. 1985). As a result, people might be less likely to permit abortion under those circumstances. In order for the abortion question order effect to occur, according to either of these explanations, respondents must consider a likely serious birth defect to be a much more compelling reason than the desire to have no more children.

Education may seem unlikely to moderate perceptual contrast effects. If one reason to obtain an abortion seems highly compelling and the other reason does not, then asking about the former should lead the latter to seem even less compelling, regardless of a respondent's cognitive skills. On the other hand, because a question order effect driven by perceptual contrast requires that a respondent does the cognitive work to recall his or her answer to a prior question and takes that answer into account when responding to the next question, less educated respondents might be less likely to manifest the question order effect because they may be less able to do that extra cognitive work.[1] If the abortion question order effect is due to subtraction, then more educated respondents might manifest a stronger question order effect because they may be more likely to have the cognitive resources available during the interview to adjust their interpretation of the question. However, Schuman and colleagues (1981) found no moderation of the order effect by education, so the same may be observed here.

Between-Country Differences

There are at least three reasons why previous findings on question order effects due to the norm of evenhandedness obtained in the Unite States may not generalize to the rest of the world. First, the norm of evenhandedness may be weaker in another country. For instance, because many Asian cultures cultivate self-concepts that are based on the principle of interdependency between people (Singelis 1994), the norm of evenhandedness appears to be much stronger in Asian countries than in Western countries (Shen, Wan, and Wyer 2011). Second, in countries with a more educated population, more people might have the cognitive capacities to take the norm of evenhandedness into account when answering the first question. Third, countries might differ from one another in the extent to which the two parties asked about in the two questions are evaluated differently. That is, if in some countries, financial contributions by unions are not evaluated differently from financial contributions by businesses, we should not expect to see the question order effect involving those groups (Schuman and Ludwig 1983).

There are also two reasons why previous findings on the abortion question order effect in the United States may not appear in the rest of the world. First, cultures may differ from one another in the extent of endorsement of the conversational norm potentially responsible for subtracting the birth defect reason from the reasons considered by the married woman (Bless and Schwarz 2010). Haberstroh and colleagues (2002) argued that respondents in collectivistic cultures think of themselves more in terms of

interdependence with others and are therefore more likely to consider what the interviewer wants to know. Accordingly, these respondents may be more likely to assume that the interviewer is not asking the same question twice, yielding stronger question order effects (Schwarz 1994). Second, there may be no perceptual contrast between the two abortion questions in a country. That is, if in some countries, abortion in case of a strong chance of a serious birth defect and in case that a married woman does not want any more children are considered equally good or bad reasons for an abortion, no question order effect should occur.

Following earlier research on cultural differences in response effects, this study explored the relations between countries' cultural characteristics and the sizes of the question order effects. Studies found, for instance, that there is less acquiescence (the tendency to agree with a statement independently of its content) in countries with a more individualistic culture compared to countries with a more collectivistic culture (He et al. 2014; Rammstedt, Danner, and Bosnjak 2017). Making use of existing databases with cultural country indicators (Hofstede 2001; Schwartz 2004), we examined whether similar associations exist with regard to question order effects.

Method

Data

Two question order experiments were implemented using 14 probability samples of the general population in 12 countries (see Table 3) as a part of the Multi-national Study of Questionnaire Design (Silber et al. 2018). Data were collected between 2013 and 2015 via the Internet in the United States (Time-Sharing Experiments for the Social Sciences (TESS) and Gallup), Canada, the Netherlands, Taiwan, Iceland, Germany (German Internet Panel), Norway, and Sweden.[2] Face-to-face interviews were conducted in Japan. A mixed-mode design was employed in Denmark (online, mail, and telephone), in Germany (GESIS Panel; online and mail), in Portugal (online and telephone), and in the United Kingdom (online and computer-assisted personal interview). The samples vary between 790 and 4,298 respondents. In total, 25,640 respondents participated in this project. A detailed description of the study setup, translation procedure, and the sampling strategy in each country is provided by Silber et al. (2018).[3] Basic methodological information for each study (Online Appendix A) and full translated question wordings for each country (Online Appendix B) can be found in the Online Appendix.

Table 3. Percentage of Cases With Missing Values in 14 Samples.

	Percentage Missing in Contributions Experiment		Percentage Missing in Abortion Experiment			
Country	Businesses	Unions	No More Children	Birth Defect	Total *N* of Cases[c]	Survey Mode[d]
United States (Gallup)	1.8	1.9	2.4	2.4	2,012	O
United States (TESS[a])	—	—	1.4	1.4	1,029	O
Canada	0.2	0.2	0.6	1.0	1,317	O
The Netherlands	1.0	0.8	0.6	0.5	2,257	O
United Kingdom	4.6	5.1	3.5	2.7	2,262	O, F2F
Denmark	1.5	1.4	1.3	1.7	1,325	O, T, M
Taiwan	0.0	0.0	0.1	0.1	790	O
Iceland	4.3	4.5	5.0	4.4	3,141	O
Germany (German Internet Panel)	0.5	0.3	0.5	0.7	1,053	O
Germany (GESIS)	2.1	2.3	2.6	2.9	4,298	O, M
Portugal[a]	—	—	0.0	0.0	1,204	O, T
Norway[b]	4.0	4.0	3.1	3.4	1,215/1,634	O
Sweden	2.9	2.9	2.9	2.9	1,770	O
Japan[a]	—	—	3.9	5.4	1,548	F2F
Total	1.9	2.0	2.3	2.3	21,440/25,640	

[a]The questions about financial contributions were not asked in the United States (TESS), Portugal, and Japan due to space limitations in the questionnaires.
[b]Different subsamples of the Norwegian Citizen Panel received the two sets of questions. A total of 1,215 respondents received the question about financial contributions, and 1,634 respondents received the questions about abortion. Only 259 respondents were asked both sets of questions.
[c]The actual sample size is larger in some country, but panel members who did not participate in the wave in which the experiments were conducted are not considered missing values. These were $n = 84$ in Germany (GIP), $n = 5$ in Norway (financial contributions), and $n = 10$ in Norway (abortion).
[d]O = online, T = telephone, M = mail, F2F = face to face.

Measures

Contributions experiment. An experiment originally implemented in a 1947 Gallup poll and first reported by Schuman and Ludwig (1983) was repeated. Respondents were randomly allocated to answer first either the question, "Do you think labor unions should be permitted to spend their money to help elect

or defeat candidates for political offices?" or the question, "Do you think businesses should be permitted to spend their money to help elect or defeat candidates for political offices?" Respondents were then asked the other question. Response options were "Yes" (coded 1) and "No" (coded 0). Due to space limitations in some questionnaires, this experiment was included in all samples except for TESS in the United States and the Portuguese and Japanese samples.

Abortion experiment. Respondents were randomly assigned to the order in which two questions about abortion were asked (Schuman and Presser 1981): "Do you think it should be possible for a pregnant woman to obtain a legal abortion if she is married and does not want any more children?" and "Do you think it should be possible for a pregnant woman to obtain a legal abortion if there is a strong chance of serious defect in the baby?" Response options were "Yes" (coded 1) and "No" (coded 0).

Education. In each country, respondents were asked about their highest level of formal education and subsequently classified as having low, medium, or high levels of formal education. The measurement and meaning of education varied across countries (see Schneider, Joye, and Wolf 2016), so experts from the GESIS methodology center in Mannheim, Germany, advised on how to best assign respondents in each country to one of the three education levels.

Analyses

Question order effects were assessed by separate χ^2 tests in each country.[4] Moderation by education was tested by estimating the parameters of logistic regression equations in each country predicting responses to the target question with an interaction between the order in which the question was asked and respondents' education along with the two main effects. Finally, all respondents were combined into one data set to test the effect of education across all samples simultaneously. For these analyses, random effects multilevel models were conducted to control for the nesting of respondents in samples (Snijders and Bosker 1999).[5]

Results

Contributions Experiment

Necessary condition. In Schuman and Ludwig's (1983) experiment, a necessary condition for the norm of evenhandedness to cause question order

Table 4. Support for Unions' and Businesses' Contributions When Asked First.

Country	Percentage Support When Asked First		Difference	χ^2	N
	Businesses	Unions			
Original result					
United States (Gallup) 1947[a]	13.3	23.2	9.9	42.88***	2,634
Meet necessary condition					
United States (Gallup)	37.0	43.4	6.4	8.50**	1,974
Canada	17.5	24.1	6.6	8.59**	1,314
The Netherlands	9.0	13.6	4.6	11.60**	2,238
United Kingdom	19.8	27.8	8.0	19.24***	2,147
Necessary condition reversed					
Denmark	46.1	33.0	−13.1	23.83***	1,317
Taiwan	20.7	9.8	−10.9	18.50***	790
Iceland	8.6	4.3	−4.3	22.85***	2,981
Do not meet necessary condition					
Germany (German Internet Panel)	14.2	13.1	−1.1	0.25	1,048
Germany (GESIS)	12.2	11.2	−1.0	0.99	4,212
Norway	20.6	25.0	4.4	3.12	1,171
Sweden	20.1	19.6	−0.5	0.05	1,719

[a]United States Gallup 1947 is the original result presented by Schuman and Ludwig (1983).
*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

effects was met: Respondents evaluated financial contributions by businesses and unions differently when those evaluations were reported first. A total of 13.3 percent of respondents said that businesses should be allowed to make financial contributions (see row 1 in Table 4), whereas 23.2 percent of respondents said that unions should be allowed to make such contributions, a highly significant difference of 9.9 percentage points ($\chi^2 = 42.88, p < .001$; Table 4).

This necessary condition was met in 7 of the 11 newly collected data sets. In four samples, businesses' financial contributions were evaluated less favorably than unions' contributions, mirroring the findings reported by Schuman and Ludwig (1983) (see rows 2–5 in Table 4). In three other samples, businesses' financial contributions received significantly more support than unions' contributions (see rows 6–8 in Table 4).

The necessary condition of more favorable evaluations of one party than the other was not met in the remaining four new data sets. Support for the right to make financial contributions was not significantly different regarding

unions and businesses (see the last four rows in Table 4). Therefore, no question order effects due to the norm of evenhandedness are expected in these four samples.

Question order effect in the United States. Schuman and Ludwig's (1983) question order effect appeared in the businesses and unions questions in the United States in 2014 (Gallup). Support for contributions by businesses increased by 5.1 percentage points when that question was preceded by the unions question ($\chi^2 = 7.85$, $p = .01$; row 2 in Table 1). And support for unions' financial contributions was significantly reduced when that question was preceded by the question about businesses ($\Delta = 5.6$ percent, $\chi^2 = 6.31$, $p = .01$).

Question order effect in other countries. In the countries in which the necessary condition for a question order effect was met, the effect appeared in all but one such country. In all of the countries in which businesses' contributions were viewed less favorably than unions' contributions (when measured initially), support for businesses' contributions increased significantly when this question was preceded by the question about unions (see rows 2–5 in Table 1). And in the countries in which businesses' contributions were viewed more favorably than unions' contributions, in all but one such country, support for union contributions increased when that question was preceded by the question about businesses (see rows 6 and 7 in Table 1).

Iceland was the only exception. Although Icelandic respondents favored financial contributions by businesses more than contributions by unions ($\Delta = 4.3$ percent, $\chi^2 = 22.85$, $p < .001$; row 8 in Table 4), support for union contributions was not affected by asking the question about businesses first ($\Delta = -0.1$ percent, $\chi^2 = .03$, $p = .98$; row 8 in Table 1). And support for financial contributions by businesses increased when that question was preceded by the question about unions ($\Delta = 4.3$ percent, $\chi^2 = 14.35$, $p < .001$).

As expected, no question order effects appeared in the countries in which contributions by unions and businesses were viewed equally favorably (last four rows in Table 1). Reversing the order in which the questions were presented did not alter the amount of support expressed for contributions by businesses or by unions.

Moderation by education. The question order effects that did appear were not consistently moderated by education. Moderation was statistically significant in the expected direction in the United States ($b = .38$, $SE = .18$, $p = .04$, Table 5). However, of the 44 tests of moderation across all countries, only

Table 5. Variation in the Contributions Question Order Effect By Education.

	Businesses Can Make Contributions				Unions Can Make Contributions			
	High Versus Medium and Low Education		Low Versus Medium and High Education		High Versus Medium and Low Education		Low Versus Medium and High Education	
Country	*z*	*p*	*z*	*p*	*z*	*p*	*z*	*p*
Meet necessary condition								
United States (Gallup)	2.06*	.04	−1.48	.14	−1.21	.23	−0.23	.82
Canada	−1.54	.12	0.05	.96	−0.05	.96	1.22	.22
The Netherlands	0.75	.46	−0.08	.94	−0.31	.76	1.17	.24
United Kingdom	0.73	.47	0.44	.66	−0.35	.73	−0.65	.52
Meet necessary condition (reversed)								
Denmark	0.57	.57	−1.00	.32	−0.08	.94	0.69	.49
Taiwan	−1.81+	.07	1.79+	.07	1.19	.23	−1.68+	.09
Iceland	1.19	.23	−0.49	.63	0.78	.44	−1.96*	.05
Do not meet necessary condition								
Germany (German Internet Panel)	1.05	.29	−1.30	.19	0.55	.58	1.33	.18
Germany (GESIS)	−0.83	.40	0.19	.85	1.60	.11	−1.15	.25
Norway	−0.14	.89	0.74	.46	0.14	.89	−0.51	.61
Sweden	0.72	.47	−0.71	.48	0.51	.61	−1.62	.11
Total sample[a]	−1.79+	0.07	−0.83	.41	−0.19	.85	−0.92	.36

Note: Z-statistics are from interaction coefficients of logistic regression models.
[a]The total effect is calculated in a random effects multilevel model.
+$p < .1$ *$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

one other yielded a statistically significant effect (4.5 percent of the tests), about what would be expected by chance alone (see Table 5). And when combining across all countries, education did not moderate the question order effect in a random effects multilevel model (see the last row in Table 5)[6]. This replicates results of Schuman and Ludwig (1983), who also did not find moderation by education.

Country characteristics. Table 6 shows that the absolute change of the support for businesses' financial contributions when that question was preceded by the unions question was strongly correlated with a number of cultural characteristics of the countries. Surprisingly, the question order effect was

Table 6. Pearson Correlations Between the Absolute Size of the Question Order Effects and Country Attributes.

	Contributions Experiment		Abortion Experiment
Country attribute	Businesses Can Make Contributions	Unions Can Make Contributions	Abortion When No More Children
Hofstede cultural dimensions[a]			
Individualism/Collectivism	.82**	.42	−.02
Power distance	.71*	−.12	−.04
Uncertainty avoidance	.05	−.33	.13
Masculinity	.42	.21	−.04
Schwartz values[b]			
Harmony	−.76*	−.47	.23
Embeddedness	.71*	.52	−.26
Hierarchy	.49	−.05	−.28
Mastery	.66*	.44	.02
Affective autonomy	−.33	−.03	.14
Intellectual autonomy	−.64*	−.53	−.17
Egalitarianism	−.61	−.38	.66*

Note: Taiwan and Iceland are not included in these analyses because country attributes were not available.
[a]Hofstede cultural dimensions are taken from Hofstede (2001).
[b]Schwartz values are taken from Schwartz (2004).
*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

stronger in countries with a more individualistic culture than in countries with a more collectivistic culture ($r = .82$, $p = .01$). Moreover, the effect size was also significantly correlated with many other characteristics, such as power distance, harmony, embeddedness, and egalitarian autonomy (Table 6). None of these correlations were statistically significant for the question about unions' financial contribution when that question was preceded by the businesses question (column 2 in Table 6). However, this may be due to the small sample size of only eight countries for which cultural country characteristics were available. In fact, the correlations between all cultural characteristics with the businesses' financial contributions effect sizes were strongly correlated with the correlations between all cultural characteristics with the unions' financial contributions effect sizes ($r = .83$, $p = .01$). This suggests that these cultural correlates of the contributions experiment were reliable.

Abortion Question Order Effect

Necessary condition. In order for the abortion question order effect to occur in a country due to perceptual contrast or subtraction, respondents must consider a birth defect to be a more compelling reason for an abortion than a married woman's desire to have no more children. This condition was met in Schuman and Presser's (1981) data: 23.3 percentage points more respondents supported permitting an abortion in the case of a risk of a serious birth defect than when a woman does not want more children ($\chi^2 = 42.88, p < .001$; see row 1 in Table 7).

This necessary condition was also met in 11 of 14 samples examined between 2013 and 2015 (rows 2–12 in Table 7). The necessary condition for the abortion question order effect was not met in the three Scandinavian countries (see the last three rows in Table 7). Equal numbers of respondents favored allowing abortion for both reasons when each question was asked first. Thus, if subtraction or perceptual contrast causes the abortion question order effect, this effect should occur only in the 11 samples in which the necessary condition was met and not in the Scandinavian countries.

Question order effect in the United States. The question order effect documented by Schuman and Presser (1981) appeared in both U.S. samples in 2014 (Gallup) and in 2014–2015 (TESS). Support for the married woman who does not want any more children getting an abortion was significantly or marginally significantly reduced by 10.2 percentage points and 5.5 percentage points, respectively, when that question was preceded by the birth defect question (Gallup, $\chi^2 = 20.96, p < .001$; TESS, $\chi^2 = 3.06, p = .08$; Table 2). As Schuman and Presser (1981) found, support for abortion in the case of a possible birth defect was not affected by question order.

Question order effect in other countries. The question order effect appeared in all but one country in which the necessary condition for that effect was met (see rows 2–12 in Table 2). That is, asking the question about a birth defect first reduced the support for abortion by the married woman who does not want any more children in all countries except Japan. In that country, as elsewhere, abortion was considered more acceptable in the case of a chance of a birth defect than for the married woman, but reversing the order of the two questions did not affect respondents' support for the married woman's right to an abortion ($\chi^2 = 0.43, p = .51$; row 12 in Table 2).

Surprisingly, the question order effect did appear in the three Scandinavian countries, even though the necessary condition for the effect to occur due to subtraction or perceptual contrast was not met in any of

Table 7. Support for Abortion when a Married Woman Does Not Want More Children and If There Is a Strong Chance of a Birth Defect when Asked First.

Country	Percentage Support When Asked First				
	Married Woman	Birth Defect	Difference	χ^2	*N*
Original result					
United States 1979[a]	60.7	84.0	23.3	42.88***	293
Meet necessary condition					
United States (Gallup)	65.0	72.9	7.9	14.45***	1,957
United States (TESS)	56.3	69.2	12.9	17.99**	1,015
Canada	80.1	88.0	7.9	15.37***	1,305
The Netherlands	72.7	91.4	18.7	133.06***	2,248
United Kingdom	76.8	89.4	12.6	62.28***	2,189
Iceland	86.2	95.1	8.9	69.72***	2,968
Germany (German Internet Panel)	80.2	91.5	11.3	27.07***	1,047
Germany (GESIS)	77.7	89.4	11.7	103.02***	4,188
Portugal	66.4	88.4	22.0	83.96***	1,204
Taiwan	77.3	94.0	16.7	44.94***	788
Japan	41.2	67.0	25.8	98.95***	1,475
Do not meet necessary condition					
Denmark	91.5	91.4	-0.1	.004	1,316
Norway	85.9	85.9	0.0	.001	1,587
Sweden	93.4	95.3	1.9	3.09+	1,721

[a]United States 1979 is the original result presented by Schuman and Presser (1981).
+$p < .1$ *$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

them (see the last three rows in Table 2). Support for abortion by the married woman who does not want any more children was reduced when that question was preceded by the birth defect question between 3.4 percentage points (Sweden, $\chi^2 = 6.64$, $p = .01$) and 10.7 percentage points (Norway, $\chi^2 = 28.66$, $p < .001$). This therefore seems to challenge the claim that the question order effect is due to subtraction or perception contrast in these or any other countries since the necessary condition proved not to be necessary.

Moderation by education. When analyzing each test separately, the question order effects that appeared were not significantly moderated by education. Of

Table 8. Variation in the Abortion Question Order Effect By Education.

	Abortion When No More Children				Abortion When Genetic Defect			
	High Versus Medium and Low Education		Low Versus Medium and High Education		High Versus Medium and Low Education		Low Versus Medium and High Education	
Country	*z*	*p*	*z*	*p*	*z*	*p*	*z*	*p*
Meet necessary condition								
United States (Gallup)	−0.44	.66	0.54	.59	0.64	.53	−0.47	.64
United States (TESS)	1.34	.18	−1.58	.12	−0.20	.84	0.20	.84
Canada	−0.92	.36	−1.52	.13	0.77	.44	0.84	.40
The Netherlands	−0.06	.95	0.62	.53	−0.16	.87	−0.45	.66
United Kingdom	0.12	.90	1.05	.30	−0.52	.60	−0.42	.68
Iceland	−0.05	.96	0.47	.64	−1.81+	.07	2.10*	.04
Germany (German Internet Panel)	0.51	.61	0.24	.81	0.48	.63	0.90	.37
Germany (GESIS)	0.57	.57	−0.18	.85	−0.91	.36	0.41	.68
Portugal	1.56	.12	−0.38	.70	−2.91**	.004	0.76	.45
Taiwan	−1.04	.30	0.39	.70	0.84	.40	0.93	.35
Japan	0.96	.34	−0.62	.53	−0.07	.95	1.56	.12
Do not meet necessary condition								
Denmark	−0.79	.43	0.89	.38	−0.78	.43	0.69	.49
Norway	1.74+	.08	−2.22*	.03	1.33	.18	−0.35	.73
Sweden	−0.60	.55	−0.16	.87	1.72+	.09	−0.34	.74
Total sample[a]	1.96*	.05	−0.36	.72	−0.43	.67	1.33	.18

Note: *Z*-statistics are for interactions in logistic regressions.
[a]The total effect is calculated in a random effects multilevel model.
+$p < .01$ *$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests).

the 56 tests, only three yielded statistically significant effects (5.4 percent of the tests), about what would be expected by chance alone (see Table 8). Only one of these three significant interactions was found for the question that showed a significant main effect (abortion when no more children). This is in line with the finding of Schuman et al. (1981), who did not find moderation by education either. Interestingly, when combining data across all samples, education moderated the abortion question order effect in a random effects multilevel model ($N = 23{,}932$; $b = .13$, $SE = .06$, $p = .05$; see the last row in Table 8)[7], suggesting that the question order effect was less strong among more educated respondents.

This pattern of data could mean that there is no moderation by education, because the effects in the individual samples were not consistently in the same direction, or it could mean that more educated respondents are less affected by question order, as suggested by the combined analyses. Either way, these findings are not in line with our prediction of a stronger question order effect among highly educated respondents.

Country characteristics. The size of the abortion question order effect did not seem to correlate with any of the available country characteristics (Table 6). The effect size correlated positively with a country's endorsement of egalitarianism ($r = .66$, $p = .03$). However, this effect was importantly driven by Japan, the only country in which the question order effect did not appear. When Japan was removed from the analysis, the correlation became nonsignificant ($r = .46$, $p = .18$). Also, none of the other correlations were significant after Japan was removed from the analysis. This suggests that the abortion question order effect is not regulated by these country-specific characteristics.

Form-Resistant Correlation Hypothesis

Finally, we used the data to test Schuman and Presser's (1981) "form-resistant correlation" hypothesis. Those investigators found that a measure correlated remarkably consistently with other variables, regardless of changes in the forming, wording, or order of the question. So, for example, answers to a question measuring an opinion correlated with the age of the respondent similarly, regardless of whether the opinion question was in one form or another.

We explored whether the rank ordering and spacing between countries in terms of their responses to a question were maintained regardless of the order in which the question was asked. And indeed, the results were striking. Treating country as the unit of analysis, the Pearson product moment correlation between answers to the business contributions question when asked first versus second (column 1 vs. column 2 in Table 1) was .97. The correlation between answers to the unions contributions question when asked first versus second (column 6 vs. column 7 in Table 1) was .95. The correlation between answers to the married woman question when asked first versus second (column 1 vs. column 2 in Table 2) was .91. And the correlation between answers to the birth defect question when asked first versus second (column 6 vs. column 7 in Table 2) was .99. Thus, researchers interested in studying cross-national differences in opinions on these issues would reach

nearly identical conclusions regardless of which question order was used to make the measurements.

Discussion

The contributions question order effect replicated in U.S. online surveys, and the effect also occurred in five of the six other countries in which respondents evaluated financial contributions by businesses and unions differently. This question order effect was not found in the four samples in which respondents did not favor financial contributions by one party over the other. These findings are therefore in line with the claim that the norm of evenhandedness causes this question order effect. The contributions question order effect was not moderated by respondent education, challenging the claim that people with strong cognitive skills are especially likely to think of the norm of evenhandedness spontaneously when asked the first question in the sequence.

The abortion question order effect replicated in two U.S. online surveys and also appeared in all but one of the other countries. As expected, the question order effect appeared in 10 of the 11 samples in which respondents considered one of the two reasons for getting an abortion to be stronger than the other (a necessary condition for perceptual contrast or subtraction). The failure of the effect to appear in the 11th country, Japan, is puzzling. Also surprising was the appearance of the question order effect in three countries in which the necessary condition was not met. This challenges both the subtraction and the perceptual contrast explanations there and perhaps in other countries as well. The question order effect was not moderated by respondent education, as expected and in line with the notion that subtraction and perceptual contrast are neither enhanced nor muted by cognitive skills.

Necessary Condition for the Norm of Evenhandedness

Results of this study suggest that researchers only need to pay attention to one of the three conditions we introduced as being necessary for a question order effect due to the norm of evenhandedness. We proposed that such question order effects only happen (1) if the norm of evenhandedness is endorsed, (2) if the parties compared in the sequence of questions are evaluated differently, and (3) if the norm of evenhandedness is not automatically activated. Reciprocal behavior seemed to be endorsed in most cultures and may be particularly strong in Asian countries (Shen et al. 2011). It is therefore likely that respondents in any survey may behave according to the norm

of evenhandedness, so researchers are well advised to assume that the first condition will always be met.

The third condition may be less useful for understanding the observed results because there was no indication of moderation of the question order effect by education. This replicates Schuman and Ludwig's (1983) finding in this regard. That is, more educated respondents were apparently not more likely to think of the norm spontaneously when answering the first question and thereby mute the magnitude of the question order effect. However, a meta-analysis of four question order experiments involving the norm of evenhandedness suggested that the question order effect was stronger among less educated respondents (see Narayan and Krosnick 1996). The present evidence challenges Narayan and Krosnick's (1996) conclusion and suggests a need to resolve this puzzle of inconsistent findings, perhaps due to changes in the relation of education to this effect over decades.

Our analyses of the association of country characteristics with the response effect may shed some light on this finding. Contrary to our expectation, the question order effect was stronger in more individualistic countries than in more collectivistic countries. This suggests that the norm of evenhandedness may be more salient in more collectivistic countries, preventing preferential treatment of the first party in a sequence of questions. Thus, the automatically activated norm of evenhandedness may prevent the occurrence of such a question order effect because respondents do not show a preference for either party compared in the sequence of questions. Accordingly, the activation of the norm of evenhandedness does not seem to depend on respondents' cognitive capacities but on the culture in which they live.

The country characteristics correlates should, however, be interpreted with some caution. This analysis was restricted to only eight countries for which cultural indicators were available (Iceland and Taiwan were not included). As a consequence, the results in the present analysis were driven by the contrast between (1) the Anglo-Saxon countries (and the Netherlands), in which respondents showed a preference for unions' contributions over that of businesses, and (2) Northern European countries in which respondents did not have such a preference. Asian countries could not be included in this analysis. Accordingly, the finding of a stronger response effect in countries that score higher on embeddedness (indicating a preference for social order, obedience, and respect for tradition, see Schwartz 2004) or mastery (more daring and ambitious), and a weaker response effect in countries with a stronger preference for harmony (unity with nature, world at peace) and for egalitarian autonomy (broadmindedness, curiosity) reflects mainly the difference between the Anglo-Saxon and Northern European countries on these

indicators. Analyses that include more countries with a broader variation on these cultural characteristics are needed to get a better understanding of the moderators of the question order effect.

Consistent with expectations, the question order effect in the contributions questions appeared only when respondents in a country had a more favorable evaluation of one of the two parties involved. In countries in which more respondents supported contributions by unions more than contributions by businesses, support for the latter increased after respondents reported support for the former. Mirroring these findings, in all but one country in which respondents favored businesses' contributions over unions' contributions, asking the unions question later increased expressed support. Furthermore, reversing the question order did not affect responses in countries in which contributions by unions and businesses received equal amounts of support. Thus, a differential evaluation of the parties involved in the comparison seems to be a necessary condition for a question order effect to appear due to the norm of evenhandedness.

Interestingly, the two-sided question order effect reported by Schuman and Ludwig (1983) appeared in the U.S. data and in Canada but in no other country in our study. That is, presenting the two questions in a sequence increased support for the less positively evaluated party and decreased support for the more positively evaluated party in the United States and Canada. In all other countries, with the exception of Iceland, presenting the questions in a sequence only increased support for the less positively evaluated party but did not affect support for the more positively evaluated party.

Perceptual Contrast in the Abortion Experiment

Most findings in the present study with regard to the abortion question order effect are in line with other research on perceptual contrast effects (Bishop et al. 1985). Earlier research has shown that people's responses to questions can be influenced if they are asked to judge similar objects on the same dimension when they have different qualitative evaluations of these objects (Tourangeau et al. 2000). For instance, politicians are evaluated more positively or negatively depending on whether the preceding question asked about a liked or disliked politician (Schwarz and Bless 1992b), and faces are evaluated as more or less attractive depending on the attractiveness of the preceding face (Wedell, Parducci, and Geisleman 1987). We found that respondents in the United States today considered not wanting any more children to be a less compelling reason for abortion than a strong chance for a serious defect in the baby. As a consequence, there was less support for

abortion by the married woman who does not want any more children when respondents were first asked about abortion when there is a chance of a birth defect than when respondents did not answer that question before. This finding is in line with earlier replications of the same experiment (Bishop et al. 1985; Schuman et al. 1981). This finding also appeared in almost all other countries examined here.

Interestingly, this question order effect also appeared in countries where the necessary condition for perceptual contrast was not observed. In three Scandinavian countries (Sweden, Norway, and Denmark), respondents expressed the highest level of support for abortion by the married woman who does not want any more children (between 86 percent and 93 percent) and an equally high level of support for abortion in case of a possibility of a birth defect when either question was asked first. In contrast to our expectations, asking the birth defect question first significantly reduced support for abortion by the married woman who does not want any more children. Also unexpected was that countries that met the necessary conditions more in Table 7 did not show a stronger question order effects in Table 2 ($r = .07$, $p = .81$).

There are two possible interpretations of these findings. One interpretation is that perceptual contrast does not underlie the abortion question order effect. However, a second interpretation is that the necessary condition was also met in the Scandinavian countries but that it was not detected in our test, because the use of only two response options was not sufficiently refined to document a difference in perception. To illustrate, imagine someone handed you a heavy stone and asked whether you think the stone is "light" or "heavy." You might pick "heavy" as your answer. Now you are handed a second, much heavier stone and have to indicate whether it is light or heavy. Given the choice between these two options, you might pick "heavy" again, even though the second stone weighs more than the first one. The dichotomous response options in the abortion experiment might have limited respondents' ability to differentiate between how compelling the provided reasons were in a similar way. Using a rating scale instead of dichotomous answer categories might allow future research to test this interpretation of our results.

The present findings challenge the subtraction account of the abortion question order effect, which has also been disputed by earlier research (Bishop et al. 1985). Subtraction requires enough cognitive capacity to recognize the relation between the two questions and to adjust the interpretation of the second question about abortion if a married woman does not want any more children by subtracting. If this occurred, the question order effect

should have been stronger among more educated respondents who had more cognitive resources available during the interview. However, just like in earlier research (Narayan and Krosnick 1996; Schuman et al. 1981), education did not moderate the effect, leading us to conclude that subtraction probably does not underlie the abortion question order effect. Moreover, the conversational norm potentially responsible for subtracting the birth defect reason from the reasons considered by the married woman is more strongly endorsed in collectivistic cultures (Bless and Schwarz 2010) and should therefore have led to a stronger question order effect in such cultures. However, our analysis did not reveal correlations between the effect sizes and countries' characteristics. These findings do not discredit the perceptual contrast explanation, which does not clearly implicate moderation by cognitive skills or country characteristics in a particular way.

Limitations

Although this is among the first studies to test for the generalizability of question order effects in multiple countries who use different languages, have different cultural backgrounds, and have different socioeconomic compositions of their populations, our investigation is restricted to only 12 countries and to only two of many question order effects (Moore 2002). The analysis of country characteristics was even more restricted, because such characteristics were not available for two countries in our sample. Testing in more countries, with more variation in perceptual contrast or contrasting evaluations of the subjects of questions asked in a sequence, and also replications of other question order effects are necessary to reach strong conclusions.

Such new research should also include tests of mode effects. Expected findings often appeared across modes, but there were some noteworthy exceptions. The abortion question order effect did not appear in Japan, which could be due to cultural reasons or due to the translation. Since abortion is up to individuals' own discretion and does not involve legal considerations in Japan, the translation was "Do you think it would be OK for a pregnant woman to abort if she is married and does not want any more children?" which could have affected responses. However, the unexpected finding in Japan could also have occurred because Japan was the only country in which data were collected only via face-to-face interviews. Mode effects could not be examined in the present research, because the samples that provided data via multiple modes did not randomly assign respondents to one mode and instead allowed respondents who did not have Internet access or did not want

to complete the questionnaire via the Internet to use a different mode (Silber et al. 2018). The failure to randomly assign respondents to modes precludes assessing the moderating effects of mode per se. Future research that includes telephone interviews in the United States would allow for direct replication of the original studies by Schuman and colleagues (1981, 1983), which involved telephone interviewing.

The wordings of some of the questions used in our experiments may have sounded dated to some respondents, even more so in some of the non-English-speaking countries after translation. This may have affected respondents' behavior but was unavoidable given the goal of this research (Silber et al. 2018). We set out to conduct well-known and often cited question order experiments that were first conducted in 1947 and 1979. Since we wanted to know whether the original question order effects would appear in the United States today and also appear in different countries, using the same question wording was unavoidable. Future research should test for similar question order effects using other questions and maybe even topics that are timelier.

Replication or Generalization?

It is interesting to compare the findings reported here to those reported by Klein et al. (2014). These investigators conducted Hyman and Sheatsley's (1950) experiment testing operation of the norm of evenhandedness in 36 separate samples, including 25 in the United States and 11 abroad (in nine countries: Brazil, the Czech Republic, Malaysia, Turkey, Canada, the United Kingdom, Poland, the Netherlands, and Italy). Hyman and Sheatsley's experiment involved asking respondents whether news reporters from communist countries should be allowed in the United States and whether American news reporters should be allowed in communist countries. In Klein et al.'s (2014) rerunning of the experiment, "communist countries" was changed to North Korea or to another country, at the discretion of the replicating investigator.

The expected question order effect was statistically significant and in the expected direction for only 36 percent of the attempted replications, and the effect was nonsignificant for 64 percent of the tests. However, combining all the samples together yielded a highly statistically significant effect in the expected direction, and a test of the homogeneity of the effects across replication attempts yielded a p value of .30, meaning that the null hypothesis of homogeneity could not be rejected. This might be interpreted as evidence of replication and generalization across countries.

The failure to observe a significant question order effect in 64 percent of the countries may have occurred because the replication attempts were vastly underpowered. Twenty-eight of the 36 samples were smaller than 150 people, meaning fewer than 75 people per cell, in contrast to the total of 1,200 participants in Hyman and Sheatsley's (1950) original study. Therefore, the test of homogeneity could also be misleading about differences between countries: Failure to reject the null hypothesis of homogeneity might be at least partly the result of lack of power as well.

Furthermore, the test assessing homogeneity might also be misleading because it was conducted across all 36 replications, 25 of which were carried out in the United States Therefore, the reported test of homogeneity is likely to be dominated by the degree of homogeneity of effects across those 25 U.S. studies since there were so few other studies. To test differences across countries, the authors could have combined all of the American respondents to compute a single effect size for the United States (and done the same for the two Canadian samples and the two samples from Poland) and then assessed the statistical significance of variation in the effect size across countries rather than across samples.

The online supplement for Klein et al.'s (2014) paper revealed that among non-U.S. countries, the question order effect was statistically significant or marginally significant in Canada, the Netherlands, the United Kingdom, Brazil, the Czech Republic, and Poland, and not even close to significant in Italy ($p = 1.0$, $N = 144$), Turkey ($p = .70$, $N = 113$), and Malaysia ($p = .46$, $N = 102$). The significant effects in Canada, the Netherlands, and the United Kingdom replicate the findings we reported on those countries. Because we collected no data in the other countries that Klein et al. (2014) investigated, we had no empirical basis for checking the robustness of their results in those places.

None of Klein et al.'s (2014) samples were drawn randomly from the populations of the nations studied, so there is no basis for expecting their results to match ours. Furthermore, even their evidence resembling Hyman and Sheatsley's (1950) is not evidence of replication, because Klein et al. (2014) did not study a random sample of American adults as Hyman and Sheatsley (1950) did. Nonetheless, it is tempting to speculate that the findings from our contributions experiment may cast some light on Klein et al.'s (2014). Perhaps their conclusion of homogeneity of effect sizes across samples is incorrect, both because so many samples were from the United States and because the samples in most countries were small. And perhaps the effect of question order is indeed absent in Italy, Turkey, and Malaysia because the necessary condition for the effect to occur was not met in those countries.

Klein et al. (2014) did not report the details of their results in a way that would allow us to identify which of their samples did and did not meet the necessary condition.

All this suggests a cautionary note about Klein et al.'s (2014) approach to interpreting their cross-national data. They did not set out to explore generalization of effects across countries and instead simply treated each implementation of an experiment outside the United States as another test of the replicability of an effect, equivalent in value to the tests done within the United States And when those authors found an instance of a nonsignificant effect, that nonsignificance was treated as a threat to the robustness of the original finding, regardless of what country the sample of respondents came from.

Our results suggest that this interpretive approach is a mistake. We showed that nonsignificant results can be explained by the failure of some countries to meet the necessary conditions. Therefore, the failure of a result to occur in some countries indicates lack of generalization rather than lack of replication. And indeed, the failures of the effect to appear in some countries actually reinforce the validity of the theory of why the question order effect occurs rather than challenging it. In other words, the theory is validated by the identification of limiting conditions. Therefore, before calling the reality of an original finding into question due to failures to replicate it across countries (see, e.g., Open Science Collaboration 2015), scholars are advised to first explore whether this variation is lack of generalization across contexts for substantive reasons rather than lack of replication. And efforts to check the robustness of a published finding may be best done with samples from the same country rather than implicitly presuming generalizability across countries.

Conclusions

Although question order effects seem to be relatively rare (Smith 1988), they appear to occur most when two questions are asked sequentially on the same topic or very similar topics (Tourangeau, Singer, and Presser 2003). We have proposed additional conditions that must be met for question order effects to occur. Thinking about such necessary conditions might help researchers to identify whether results of an experiment are attributable to a proposed mechanism. We look forward to more research seeking to explain response effects like these. In the meantime, the present study suggests that international research may be fruitful for assessing the replicability, generalizability, and mechanisms of question design effects.

Authors' Note

Data used in this article can be downloaded from the study website https://osf.io/gwrve/.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article uses data from the German Internet Panel waves 8 (doi: 10.4232/1.12614) and 9 (doi: 10.4232/1.12615). A study description can be found in Blom, Gathmann, and Krieger (2015). The German Internet Panel is the central data collection (project Z1) of Collaborative Research Center 884 "Political Economy of Reforms" (SFB 884) at the University of Mannheim and is funded by the German Research Foundation (DFG). The Longitudinal Internet Studies for the Social Sciences (LISS) Panel data were collected by CentERdata (Tilburg University, the Netherlands) through its Measurement and Experimentation in the Social Sciences (MESS) project funded by the Netherlands Organization for Scientific Research (grant number 176.010.2005.017). This article uses data from the GESIS Panel (doi: 10.4232/1.12658). The development of the GESIS Panel was funded the German Federal Ministry of Education and Research and is now a permanent data collection facility operated by GESIS. The SSRI Online Panel data were collected by the Social Science Research Institute, University of Iceland. The panel is funded by the institute. The Understanding Society Innovation Panel is funded by the UK Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service: Understanding Society: Innovation Panel, Waves 1–8, 2008–2015 [data collection]. 7th Edition. UK Data Service. SN: 6849. The data in Sweden were collected by the Laboratory of Opinion Research at the University of Gothenburg through its Swedish Citizen Panel. This data collection is described in Martinsson et al. (2014). The Danish data are funded by The Danish Council for Independent Research. The Norwegian Citizen Panel/DIGSSCORE is funded by the University of Bergen, Uni Research, and the Bergen Research Foundation (grant no. 01234). The data of Taiwan were collected under the funding of Research Center for Humanities and Social Sciences, Academia Sinica. The Japan survey was funded by the Environment Research and Technology Development Fund (1-1406) of the Ministry of the Environment, Japan. The data collection for Portugal has benefited from funding of the Portuguese Foundation for Science and Technology (FCT), grant number PTDC/

IVC-CPO/3921/2012. TESS (Time-Sharing Experiments for the Social Sciences) is funded by the National Science Foundation (SES-1628057).

ORCID iD

Tobias H. Stark https://orcid.org/0000-0002-3163-5776
Henning Silber https://orcid.org/0000-0002-3568-3257
Jon A. Krosnick https://orcid.org/0000-0003-0336-1613
Ana Belchior https://orcid.org/0000-0003-3387-2529
Michael Bosnjak https://orcid.org/0000-0002-1431-8461
Sanne Lund Clement https://orcid.org/0000-0002-5480-7622
Guðbjörg Andrea Jónsdóttir https://orcid.org/0000-0002-8852-9560
Peter Lynn https://orcid.org/0000-0002-3493-5686
Ditte Shamshiri-Petersen https://orcid.org/0000-0003-2766-0220
Endre Tvinnereim https://orcid.org/0000-0001-8891-2503

Supplemental Material

Supplemental material for this article is available online.

Notes

1. The abortion question order effect has been found to be stronger among younger respondents, perhaps because young respondents have better short-term memory capacity, allowing better recollection of a preceding question when answering the subsequent one, a necessary condition for perceptual contrast or subtraction (Knäuper et al. 2007).
2. Data were also collected in France as part of this project, but the randomization of the experiments was not implemented correctly, making it impossible to explore question order effects.
3. To minimize potential problems due to translation errors, the source questionnaire was provided in English and teams in the various countries developed functionally equivalent translations following the Translation, Review, Adjudication, Pretesting, and Documentation approach (Harkness, Villar, and Edwards 2010).
4. In order to maximize sample representativeness, it would be desirable to use poststratification weights for each sample. When the data collections were initiated, we assumed that we would be able to produce poststratification weights. However, too little information about population statistics for multiple country was available to generate proper weights. Furthermore, we received weights only for 7 of the 14 samples in this study, and one set of weights from the United States did not reproduce the nation's demographics. We therefore report unweighted results in this article and report weighted results in Online Appendix C, but we caution readers not to assume that the weighted results are optimal.

5. There were relatively few missing answers to the four target questions (see Table 3), so we removed the few missing cases.
6. Of the 20,922 respondents who gave valid answers to the business questions, 412 (1.97 percent) had missing values on the education variable, and for 37 respondents from Iceland, the order in which the business/union questions were asked was not recorded, so these respondents were excluded from the moderation analyses. Of the 20,905 respondents who gave valid answers to the union questions, 411 (1.97 percent) had missing values on the education variable and were removed from the sample.
7. Of the 25,023 respondents with valid answers on the abortion when a women does not want more children question, 1,075 (4.30 percent) had missing values on the education variable, and for 15 cases in the Japanese data, the order in which the abortion questions were asked was not recorded, so these cases were not included in the analyses. Of the 25,006 respondents with valid answers to the birth defect question, 1,069 (4.27 percent) has missing values on education and for 66 cases the order in which the questions were asked was not recorded.

References

Bishop, George F., Robert W. Oldendick, and Alfred J. Tuchfarber. 1985. "The Importance of Replicating a Failure to Replicate—Order Effects on Abortion Items." *Public Opinion Quarterly* 49:105-14.

Bless, Herbert and Norbert Schwarz. 2010. "Mental Construal and the Emergence of Assimilation and Contrast Effects: The Inclusion/Exclusion Model." *Advances in Experimental Social Psychology* 42:319-73.

Blom, Annelies G., Christina Gathmann, and Ulrich Krieger. 2015. "Setting Up an Online Panel Representative of the General Population: The German Internet Panel." *Field Methods* 27:391-408.

Bollen, Kenneth, John T. Cacioppo, Robert Kaplan, Jon A. Krosnick, and James L. Olds. 2015. "Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science." Pp. 1-29 in *Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences*. Arlington, Virginia: National Science Foundation. Retrieved (02/21/2017) from: https://www.nsf.gov/sbe/AC_Materials/SBE_Robust_and_Reliable_Research_Report.pdf

Grice, H. Paul. 1975. "Logic and Conversation." Pp. 41-58 in *Syntax and Semantics*, edited by P. Cole and J. L. Morgan. New York: Academic Press.

Haberstroh, Susanne, Daphna Oyserman, Norbert Schwarz, Ulrich Kühnen, and Li-Jun Ji. 2002. "Is the Interdependent Self More Sensitive to Question Context than the Independent Self? Self-construal and the Observation of Conversational Norms." *Journal of Experimental Social Psychology* 38:323-29.

Harkness, Janet A., Ana Villar, and Brad Edwards. 2010. "Translation, Adaptation, and Design." Pp. 115-140 in *Survey Methods in Multinational, Multiregional, and Multicultural Contexts*, edited by Janet A. Harkness, Michael Braun, Brad Edwards, Timothy P. Johnson, Lars E. Lyberg, Peter Ph. Mohler, Beth-Ellen Pennell, and Tom W. Smith. New Jersey: Wiley.

He, Jia, Alejandra D. C. Dominguez Espinosa, Ype H. Poortinga, and Fons J. R. van de Vijver. 2014. "Acquiescent and Socially Desirable Response Styles in Cross-cultural Value Surveys." Pp. 98-111 in *Toward Sustainable Development through Nurturing Diversity*, edited by Leon T. B. Jackson, Deon Meiring, Fons J. R. van de Vijver, and Erdhabor Idemudia. Melbourne, FL: International Association for Cross-Cultural Psychology.

Higgins, E. Tory and Liora Lurie. 1957. "Context, Categorization, and Recall: The "Change-of-standard" Effect." *Cognitive Psychology* 15:525-47.

Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations*. Thousand Oaks, CA: Sage.

Hyman, Herbert H. and Paul B. Sheatsley. 1950. "The Current Status of American Public Opinion." Pp. 11-34 in *The Teaching of Contemporary Affairs*, Twenty-first Yearbook of the National Council of Social Studies, edited by J. C. Payne. Washington, DC: National Council of Social Studies.

Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams, Štěpán Bahník, Michael J. Bernstein, Konrad Bocian, Mark J. Brandt, Beach Brooks, Claudia Chloe Brumbaugh, Zeynep Cemalcilar, Jesse Chandler, Winnee Cheong, William E. Davis, Thierry Devos, Matthew Eisner, Natalia Frankowska, David Furrow, Elisa Maria Galliani, Fred Hasselman, Joshua A. Hicks, James F. Hovermale, S. Jane Hunt, Jeffrey R. Huntsinger, Hans Ijzerman, Melissa Sue John, Jennifer A. Joy-Gaba, Heather Barry Kappes, Lacy E. Krueger, Jaime Kurtz, Carmel A. Levitan, Robyn K. Mallett, Wendy L. Morris, Anthony J. Nelson, Jason A. Nier, Grant Packard, Ronaldo Pilati, Abraham M. Rutchick, Kathleen Schmidt, Jeanine L. Skorinko, Robert Smith, Troy G. Steiner, Justin Storbeck, Lyn M. Van Swol, Donna Thompson, A. E. Van 'T Veer, Leigh Ann Vaughn, Marek Vranka, Aaron L. Wichman, Julie A. Woodzicka, and Brian A. Nosek. 2014. "Investigating Variation in Replicability A "Many Labs" Replication Project." *Social Psychology* 45:142-52.

Knäuper, Bärbel, Norbert Schwarz, Denise Park, and Andreas Fritsch. 2007. "The Perils of Interpreting Age Differences in Attitude Reports: Question Order Effects Decrease with Age." *Journal of Official Statistics* 23:515-28.

Krosnick, Jon A. and Stanley Presser. 2010. "Questionnaire Design." Pp. 263-314 in *Handbook of Survey Research*, edited by J. D. Wright and P. V. Marsden. West Yorkshire, UK: Emerald Group.

McFarland, Sam G. 1981. "Effects of Question Order on Survey Responses." *Public Opinion Quarterly* 45:208-15.

Moore, David W. 2002. "Measuring New Types of Question-order Effects—Additive and Subtractive." *Public Opinion Quarterly* 66:80-91.

Narayan, Sowmya and Jon A. Krosnick. 1996. "Education Moderates Some Response Effects in Attitude Measurement." *Public Opinion Quarterly* 60:58-88.

Nie, Norman H., Saar Golde, and Daniel M. Butler. 2009. "Education and Verbal Ability Over Time: Evidence from Three Multi-time Sources." Accessed January 12, 2017. Retrieved from (http://www.learningace.com/doc/1391311/129f7bd3448db86d19dba0cae2553072/education_and_verbal_ability).

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." Pp. aac4716 in *Science* 349.

Rammstedt, Beatrice, Daniel Danner, and Michael Bosnjak. 2017. "Acquiescence Response Styles: A Multilevel Model Explaining Individual-level and Country-level Differences." *Personality and Individual Differences* 107:190-94.

Saris, Willem E. and Irmtraud N. Gallhofer. 2011. "The Results of the MTMM Experiments in Round 2." RECSM Working Paper Series, Working Paper No. 23, Universtiat Pompeu Fabra, Barcelona, Spain.

Saris, Willem E. and Irmtraud N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. Hoboken, NJ: Wiley.

Schneider, Silke, Dominique Joye, and Christof Wolf. 2016. "When Translation Is Not Enough: Background Variables in Comparative Surveys."Pp. 288-307 in *The SAGE Handbook of Survey Methodology*, edited by Christof Wolf, Dominique Joye, Tom W. Smith, and Yang-chih Fu. London, UK: Sage.

Schuman, Howard and Jacob Ludwig. 1983. "The Norm of Even-handedness in Surveys as in Life." *American Sociological Review* 48:112-20.

Schuman, Howard and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.

Schuman, Howard, Stanley Presser, and Jacob Ludwig. 1981. "Context Effects on Survey Responses to Questions About Abortion." *Public Opinion Quarterly* 45: 216-23.

Schwartz, Shalom H. 2004. "Mapping and Interpreting Cultural Differences Around the World." Pp. 43-73 in *Comparing Cultures: Dimensions of Culture in a Comparative Perspective*, edited by Henk Vinken, Jospeh Soeters, and Peter Ester. Leiden, the Netherlands: Brill.

Schwarz, Norbert. 1994. "Judgment in Social Context: Biases, Shortcomings, and the Logic of Conversation." Pp. 123-62 in *Advances in Experimental Psychology*, edited by M. P. Zanna. San Diego, CA: Academic Press.

Schwarz, Norbert and Herbert Bless. 1992a. "Constructing Reality and Its Alternatives: An Inclusion/Exclusion Model of Assimilation and Contrast Effects in Social Judgment." Pp. 217-45 in *The Construction of Social Judgment*. edited by Leonard L. Martin and Abraham Tesser. Hillsdale, NJ: Erlbaum.

Schwarz, Norbert and Herbert Bless. 1992b. "Scandals and Public Trust in Politicians: Assimilation and Contrast Effects." *Personality and Social Psychology Bulletin* 18:574-79.

Shen, Hao, Fang Wan, and Robert S. Wyer. 2011. "Cross-cultural Differences in the Refusal to Accept a Small Gift: The Differential Influence of Reciprocity Norms on Asians and North Americans." *Journal of Personality and Social Psychology* 100:271-81.

Silber, Henning, Tobias H. Stark, Annelies G. Blom, and Jon A. Krosnick. 2018. "Implementing a Multi-National Study of Questionnaire Design." in *Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC)*, edited by T. P. Johnson, B.-E. Pennell, I. Stoop, and B. Dorer. New York, NY: Wiley.

Singelis, Theodore M. 1994. "The Measurement of Independent and Interdependent Self-construals." *Personality and Social Psychology Bulletin* 20:580-91.

Smith, Tom W. 1988. *Ballot Position: An Analysis of Context Effects Related to Rotation Design*. GSS Methodological Report No. 55. Chicago, IL: National Opinion Research Center.

Snijders, Tom A. B. and Roel J. Bosker. 1999. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London, UK: Sage.

Stevens, Stanley S. 1957. "On the Psychophysical Law." *The Psychological Review* 64:153-81.

Tourangeau, Roger and K. A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Pychological Bulletin* 103:299-314.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, MA: Cambridge University Press.

Tourangeau, Roger, Eleanor Singer, and Stanley Presser. 2003. "Context Effects in Attitude Surveys: Effects of Remote Items and Impact on Predictive Validity." *Sociological Methods & Research* 31:486-513.

Wedell, Douglas H., Allen Parducci, and R. Edward Geisleman. 1987. "A Formal Analysis of Ratings of Physical Attractiveness: Successive Contrast and Simultaneous Assimilation." *Journal of Experimental Social Psychology* 23:230-49.

Author Biographies

Tobias H. Stark is assistant professor in the department of Interdisciplinary Social Science at Utrecht University, the Netherlands. His research focuses on the intersection of prejudice and interethnic relations in social networks. His full CV is at www.tobiasstark.nl

Henning Silber (GESIS - Leibniz Institute for the Social Sciences, Department of Survey Design and Methodology, Mannheim, Germany) is head of the Survey

Operations team at GESIS. His research interests include international survey methodology, political sociology, and the experimental social sciences.

Jon A. Krosnick is Frederick O. Glover Professor in Humanities and Social Sciences, and Professor of Communication, Political Science, and Psychology at Stanford University, Director of Stanford's Political Psychology Research Group, and Research Psychologist at the U.S. Census Bureau.

Annelies G. Blom is a Professor for Methods of empirical Social Research at the Department of Political Science, University of Mannheim (Germany) and the Principal Investigator of the German Internet Panel.

Midori Aoyagi is a Principle Researcher, Center for Social and Environmental Systems, National Institute for Environmental Studies (NIES), Tsukuba, Japan. Her research interests are public perception and support for climate change policies, and media effects on this issue. Also interested in sustainable consumption, lifestyles. She graduated from Kyoto University.

Ana Belchior is an assistant professor, with aggregation, in the Department of Political Science and Public Policies at ISCTE-IUL (Instituto Universitário de Lisboa), and researcher at CIES-IUL. She is currently the head of the Department of Political Science and Public Policies and the coordinator of the Master in Political Science, both at ISCTE-IUL (curriculum at: https://ciencia.iscte-iul.pt/authors/ana-maria-da-conceicao-belchior/cv)

Michael Bosnjak is director of ZPID - Leibniz Institute for Psychology Information in Trier, Germany, and Full Professor of Psychological Research Methods at the University of Trier. His research interests include: Research synthesis methods, consumer/business psychology, and survey methodology.

Sanne Lund Clement is Associate Professor at the Department of Political Science, Aalborg University, Denmark and leader of the Political Sociology Research Group at Aalborg University. Her research field includes survey research and survey quality, public opinion and political behavior.

Melvin John has majored in political science at the University of Bremen and in psychology at the University of Mannheim. Currently he is studying to receive his masters in psychology at the University of Mannheim.

Guðbjörg Andrea Jónsdóttir (Ph.D. Social Psychology, 2004 from London School of Economics and Political Science) is Director of the Social Science Research Institute at the University of Iceland. Prior to that she was Research Director with Gallup in Iceland. Her main research interests are methodology and public opinion, in particular context and order effects in social surveys, wording of questions, mode effects and response rate.

Karen Lawson (PhD, CE) is a Professor of Applied Social Psychology at the University of Saskatchewan, Canada. Her program of research focuses on reproductive psychology and program evaluation theory, methodology, and training.

Peter Lynn is Professor of Survey Methodology at the University of Essex, where he specialises in all aspects of survey design and data collection. He is co-investigator of Understanding Society: the UK Household Longitudinal Study and is currently President of the International Association of Survey Statisticians. His full CV is at www.iser.essex.ac.uk/people/plynn.

Johan Martinsson is associate professor of Political Science at the University of Gothenburg, Sweden, and research director of the SOM Institute. His research is focused on voting behaviour, environmental attitudes and survey methodology.

Ditte Shamshiri-Petersen is associate professor at Department of Political Science, Aalborg University, Denmark. Her research field includes political behavior; particularly electoral behavior, opinion formation and public opinion, as well as methodological research.

Endre Tvinnereim is Senior Researcher at the Uni Research Rokkan Centre for Social Studies and the Centre for Climate and Energy Transformation (CET) at the University of Bergen. His main research interests are cross-national public opinion on climate policies, quantitative text analysis in survey research, and evidence-based evaluations of cap-and-trade and other CO2 pricing mechanisms. Tvinnereim co-chairs the Climate and Environment research group of the Norwegian Citizen Panel.

Ruoh-rong Yu is now a Research Fellow and the Executive Director of the Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, Taiwan. Her research interests include survey methodology, family studies, and labor research. She has been the PI of the Taiwan Panel Study of Family Dynamics Project since 2013.