



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Lack of replication or generalization? Cultural values explain a question wording effect

Silber, Henning; Tvinnereim, Endre ; Stark , Tobias; Blom, Annelies G.; Krosnick, Jon A.; Bosnjak, Michael ; Clement, Sanne Lund; Cornilleau, Anne ; Cousteaux, Anne-Sophie ; John, Melvin ; Jónsdóttir, Guðbjörg Andrea ; Lawson, Karen ; Lynn, Peter ; Martinsson, Johan ; Shamshiri-Petersen, Ditte; Tu, Su-Hao

Published in:

Journal of Survey Statistics and Methodology

DOI (link to publication from Publisher):

[10.1093/jssam/smab007](https://doi.org/10.1093/jssam/smab007)

Publication date:

2022

Document Version

Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Silber, H., Tvinnereim, E., Stark, T., Blom, A. G., Krosnick, J. A., Bosnjak, M., Clement, S. L., Cornilleau, A., Cousteaux, A.-S., John, M., Jónsdóttir, G. A., Lawson, K., Lynn, P., Martinsson, J., Shamshiri-Petersen, D., & Tu, S.-H. (2022). Lack of replication or generalization? Cultural values explain a question wording effect. *Journal of Survey Statistics and Methodology*, 10(5), 1121–1147. <https://doi.org/10.1093/jssam/smab007>

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 -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: February 06, 2025

**Lack of Replication or Generalization?
Cultural Values Explain a Question Wording Effect**

Henning Silber
GESIS – Leibniz Institute for the Social
Sciences

Anne-Sophie Cousteaux
Sciences Po

Endre Tvinnereim
University of Bergen/NORCE

Melvin John
University of Mannheim

Tobias H. Stark
Utrecht University/ICS

Guðbjörg Andrea Jónsdóttir
University of Iceland

Annelies G. Blom
University of Mannheim

Karen Lawson
University of Saskatchewan

Jon A. Krosnick
Stanford University

Peter Lynn
University of Essex

Michael Bosnjak
Leibniz Institute for Psychology Information

Johan Martinsson
University of Gothenburg

Sanne Lund Clement
Aalborg University

Ditte Shamshiri-Petersen
Aalborg University

Anne Cornilleau
Sciences Po

Su-hao Tu
Academia Sinica

This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Journal of Survey Statistics and Methodology* following peer review. The version of record (published version) [Lack of Replication or Generalization? Cultural Values Explain a Question Wording Effect; Henning Silber et al; Journal of Survey Statistics and Methodology 2021; doi: 10.1093/jssam/smab007] is available online at:

<https://academic.oup.com/jssam/advance-article/doi/10.1093/jssam/smab007/6355346>

and should be cited as follows:

Silber, Henning et al (2021) 'Lack of Replication or Generalization? Cultural Values Explain a Question Wording Effect'. *Journal of Survey Statistics and Methodology*, published online 20 August 2021.

Author Contact Information

Henning Silber, Department of Survey Design and Methodology, GESIS – Leibniz-Institute for Social Sciences, B2 1, 68159 Mannheim, Germany, email: henning.silber@gesis.org.

Endre Tvinnereim, University of Bergen, Department of Administration and Organization Theory and NORCE Norwegian Research Centre AS, Box 7802, 5020 Bergen, Norway, e-mail: endre.tvinnereim@uib.no.

Tobias H. Stark, ERCOMER (European Research Centre on Migration and Ethnic Relations), Utrecht University/ICS, Padualaan 14, 3573AT Utrecht, The Netherlands, email: t.h.stark@uu.nl.

Annelies G. Blom, Department of Political Science, School of Social Sciences, and Collaborative Research Center 884 "Political Economy of Reforms", University of Mannheim, B6 30-32, 68131 Mannheim, Germany, email: blom@uni-mannheim.de

Jon A. Krosnick, Departments of Communication, Political Science, and Psychology, Stanford University, Stanford, CA. 450 Serra Mall, Stanford, CA 94305, USA of America, email: krosnick@stanford.edu. Jon Krosnick is University Fellow at Resources for the Future.

Michael Bosnjak, Leibniz Institute for Psychology Information (ZPID), Universitätsring 15, 54296 Trier, Germany, email: mb@leibniz-psychology.org.

Sanne Lund Clement, Department of Political Science, Aalborg University, Fibigerstraede 1, 9220 Aalborg East, Denmark, e-mail: clement@dps.aau.dk.

Anne Cornilleau, Centre de données socio-politiques (Sciences Po/CNRS), 27 rue Saint Guillaume 75337 Paris cedex, France, email : anne.cornilleau@sciencespo.fr.

Sanne Lund Clement, Department of Political Science, Aalborg University, Fibigerstraede 1, 9220 Aalborg East, Denmark, e-mail: clement@dps.aau.dk.

Melvin John, Collaborative Research Center 884 "Political Economy of Reforms", University of Mannheim, L13 15-17, 68131 Mannheim, Germany, email: mjohn@mail.uni-mannheim.de

Guðbjörg Andrea Jónsdóttir, Social Science Research Institute, University of Iceland, Sæmundargata 10, 101 Reykjavík, Iceland, email: gudbjorg@hi.is.

Karen Lawson, Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, SK, S7N5A5, Canada, email: karen.lawson@usask.ca.

Peter Lynn, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, email: plynn@essex.ac.uk.

Johan Martinsson, Department of Political Science, University of Gothenburg, Box 100, 405 30 Gothenburg, Sweden, email: johan.martinsson@pol.gu.se.

Ditte Shamshiri-Petersen, Department of Political Science, Aalborg University, Fibigerstraede 1, 9220 Aalborg East, Denmark, e-mail: dittep@dps.aau.dk.

Su-hao Tu, Center for Survey Research, Research Center for Humanities and Social Sciences, Academia Sinica, 128 Section 2 Academia Road, Nankang 11529, Taipei, Taiwan, email: suhao@gate.sinica.edu.tw

Author Notes

Gallup collected data in the U.S. from their online panel and funded that data collection.

Data were also collected in the U.S. through the project entitled Time-sharing Experiments for the Social Sciences (TESS), funded by the National Science Foundation (SES-1628057).

Data from Germany were collected from the German Internet Panel (waves 8 (DOI: 10.4232/1.12614) and 9 (DOI: 10.4232/1.12615)), funded by the German Research Foundation through the Collaborative Research Center (SBF) 884 "Political Economy of Reforms" (Project-ID: 139943784). A study description can be found in Blom et al. (2015).

Other data from Germany were collected from the GESIS Panel (DOI: 10.4232/1.12658), funded by the German Federal Ministry of Education and Research and GESIS.

The data from Taiwan were collected using funding from the Research Center for Humanities and Social Sciences, Academia Sinica.

Data from the Netherlands were collected from the LISS Panel by CentERdata (Tilburg University, The Netherlands) through its MESS project, which was funded by the Netherlands Organization for Scientific Research (grant number 176.010.2005.017).

The College of Arts and Sciences at the University of Saskatchewan funded the Canadian data collection.

The Norwegian Citizen Panel/DIGSSCORE is funded by the University of Bergen, NORCE, and the Bergen Research Foundation (grant no. 01234).

The data from Sweden were collected by the Laboratory of Opinion Research (LORE) at the University of Gothenburg through its Swedish Citizen Panel (SCP). This data collection is described in Martinsson et al. (2014).

The Danish data were collected by the University of Aalborg, funded by the Danish Council for Independent Research.

The Icelandic data were collected by the Social Science Research Institute via the SSRI Online Panel, University of Iceland. The panel is funded by the Institute.

UK data were collected from the Understanding Society Innovation Panel, 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. Survey data collection was managed 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 from France were collected from the ELIPSS panel, which is part of the project entitled Data, Infrastructure, Methods of Investigation in the Social Sciences and Humanities (DIME-SHS) and supported by a grant from the French National Research Agency (ANR-10-EQPX-19-01).

The authors thank David S. Yeager for his comments on an earlier version of this manuscript.

**Lack of Replication or Generalization?
Cultural Values Explain a Question Wording Effect**

Abstract

In the context of the current “replication crisis” across the sciences, failures to reproduce a finding are often viewed as discrediting it. This paper shows how such a conclusion can be incorrect. Schuman and Presser (1981) showed that including the word “freedom” in a survey question significantly increased approval of allowing a speech against religion in the U.S. New experiments in probability sample surveys (N = 23,370) in the U.S. and 10 other countries showed that the wording effect replicated in the U.S. and appeared in four other countries (Canada, Germany, Taiwan, and the Netherlands) but not in the remaining countries. The effect appeared only in countries in which the value of freedom is especially salient and endorsed. Thus, public support for a proposition was enhanced by portraying it as embodying a salient principle of a nation’s culture. Instead of questioning initial findings, inconsistent results across countries signal limits on generalizability and identify an important moderator.

Keywords

Replication, generalization, survey research methods, questionnaire design, question wording effects, cross-cultural comparisons

During the last decade, the “crisis of replicability” has exploded across the sciences, including psychology (Pashler and Wagenmakers 2012), economics (Camerer et al. 2016), medicine (Ioannidis 2005), and other fields (Baker 2016). Failures to replicate highly visible and often cited effects (Open Science Collaboration et al. 2015), evidence of the so-called “decline effect”, whereby an initially strong experimental treatment effect becomes weaker and weaker across subsequent tests of it (Schooler 2011), and other findings have led to wide-spread questioning of how reliable the published scientific literatures are. And documentation of strategies that researchers have implemented that create illusory findings (John et al. 2012; Simmons et al. 2011) have reinforced readers’ uncertainty about whether they can count on the validity of authors’ descriptions, inferences, and conclusions.

In efforts to gauge the magnitude of the problem, some teams of investigators have conducted systematic investigations of the replicability of published empirical findings. For example, multiple publications have now reported the results obtained when a large number of international teams of researchers have conducted the same psychological experiment in various different countries (e.g., Camerer et al. 2018; Ebersole et al. 2016; Klein et al. 2014). These publications quickly became highly cited (e.g., Klein et al. 2014, had been cited in 428 publications as of September 1, 2018 according to Google Scholar) and have even found their way onto the front page of the New York Times (Carey 2015). The headline of the printed edition of that newspaper story focused on the fact that most past studies could not be replicated.

Some of the experiments examined in those studies involved manipulations of survey question wording and question order, and perhaps surprisingly, the results of those experiments turned out to be unusually robust. One such effect involved a question wording experiment comparing opinions about whether speeches against democracy should be forbidden or should be allowed. In the first study of the phenomenon in the U.S. conducted by Rugg (1941), more

respondents said that such speeches should not be allowed than said they should be forbidden. The same wording effect was observed in 36 replications conducted in ten countries in seven languages, thus attesting to the robustness of the effect (Klein et al. 2014).

A somewhat different but nonetheless reassuring pattern of findings was obtained by Klein et al. (2014) when they attempted to replicate an experiment manipulating question order. In 1948, Hyman and Sheatsley (1950) documented the operation of the norm of even-handedness: in an American sample, 37% of respondents said that communist news reporters should be allowed into the U.S. to report information to their home countries, whereas 73% of other respondents expressed this opinion after first being asked whether a communist country like Russia should admit American news reporters (Hyman and Sheatsley 1950). Klein et al. (2014) conducted that experiment in 36 separate samples, including 25 in the U.S. and 11 in nine other countries: Brazil, the Czech Republic, Malaysia, Turkey, Canada, the United Kingdom, Poland, the Netherlands, and Italy.¹

The expected question order effect was statistically significant and in the expected direction for only 36% of the attempted replications, and the effect was non-significant for 64% 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 samples yielded a p-value of .30, meaning that the null hypothesis of homogeneity could not be rejected, again attesting to the robustness of the effect.

These findings of robustness are important, because Rugg's (1941) and Hyman and Sheatsley's (1950) experiments have ended up being archetypes for thousands of such

¹ Hyman and Sheatsley's (1950) experiment involved asking respondents whether news reporters from communist countries should be allowed in the U.S. and whether American news reporters should be allowed in communist countries. In Klein et al.'s (2014) re-running of the experiment, "communist countries" was changed to North Korea or to another country, at the discretion of the replicating investigator.

experiments examining whether changing the form, wording, and context of a survey question (often in subtle ways that at first seem innocuous) alter the distributions of responses obtained. And such experiments have demonstrated numerous effects of such manipulations (e.g., Schuman and Presser 1981).

Although most of the published evidence on such effects comes from studies conducted with American participants using English-language questionnaires, some questionnaire design effects have been studied cross-nationally in the European Social Survey during the last decade (e.g., Saris and Gallhofer 2007), and an increasing number of one-off papers have explored response effects in individual countries. But even today, the vast majority of evidence on question design effects comes from studies in the U.S., many done decades ago. And yet such evidence often guides questionnaire design across nations, despite the absence of evidence of the generalizability of those effects across countries and languages. This is why Klein et al.'s (2014) replications of question wording and order effects across many samples in the U.S. and in other countries is reassuring and valuable.

In this paper, we report similar tests of another question wording effect documented initially by Schuman and Presser (1981). Unlike Klein et al. (2014), whose investigators collected data from haphazard, convenience samples, we collected data from probability samples of the populations of many countries, including multiple samples within two countries. These data allowed us to gauge replicability of the effect within country and to gauge generalizability of the effect across countries in a more rigorous way (for evidence on the importance of probability sampling for survey accuracy, see Cornesse et al. (2020), and for reporting precision, see Cornesse and Blom (2020)). And the results we describe are strikingly different from those reported by Klein et al. (2014), regarding the robustness of question design effects.

Question Wording Experiment

Schuman and Presser's (1981) experiment involved altering the wording of a question about whether a person should be allowed to speak publicly against churches and religion, asked in the following two ways:

- (1) There are some people who are against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed to speak, or not?
- (2) There are some people who are against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed **the freedom** to speak, or not?

The only difference between these two versions of the question is the omission vs. inclusion of the words "the freedom". The two versions of the question communicate the same information and seek the same decision from respondents. But the second version highlights the implication that if a person is not allowed to speak, he or she is deprived of freedom.

The impact of this wording manipulation seems likely to depend upon the importance that a respondent attaches to the value of freedom. Specifically, adding "freedom" to the question seems likely to increase willingness to allow the speech only among respondents who consider freedom to be an important value. People who do not value freedom should be uninfluenced by addition of the word.

This logic anticipates a strong effect of adding that word to the question if the experiment is conducted in the U.S., where Schuman and Presser (1981) did it, because Americans seem to attach great importance to this value. For example, the American national anthem describes the country as "the land of the free". The national anthem is sung often by Americans at the beginning of sports events and other sorts of events that millions of Americans witness numerous

times each year. And when performed, singers sustain the word “free” under a prolonged *fermata*, thus emphasizing it. Furthermore, the U.S. Constitution talks about “free persons”, securing “freedom and happiness”, “freedom of expression”, and “freedom of speech”. Thus, freedom is a term with special meaning and appeal to Americans.

This has two useful implications in the context of the present investigation. First, the presence and emphasis of “free” in the national anthem and the Constitution may signal great cultural significance of the value. And second, repeated exposure of most Americans to the word during countless performances of the anthem constitutes cognitive priming of the concept (Bargh et al. 2001; Weingarten et al. 2016), which itself will enhance its presence in people’s thinking. Thus, it might seem likely that associating the notion of a speech against churches and religions in Schuman and Presser’s (1981) experiment with the idea of freedom would increase Americans’ willingness to permit such a speech.

That is exactly what Schuman and Presser (1981, pp. 290–291) found. In an experiment conducted in a telephone survey in 1974, members of a probability sample of American adults were randomly assigned to hear one of the two question wordings, omitting or including the word “freedom.” Of respondents asked the question without “freedom,” 65.8% said the speech should be permitted. And of respondents who heard the word “freedom,” this percentage rose to 71.7%, a statistically significant increase of 5.9 percentage points (Table 11.8, p. 291). In the study described in this paper, we explored whether this effect appeared in the U. S. forty years later and whether it appeared in other countries as well. Thus, we tested replicability and generalizability.

Generalizability of the Question Wording Effect

Should we expect such generalization across countries? One possibility is that freedom is universally valued by contemporary adults world-wide, so framing a speech in terms of freedom should yield the same increase in support seen by Schuman and Presser (1981) in the U.S. But it

is also possible that nations differ in the extent to which they emphasize the value of freedom and the frequency with which rituals prime the concept in the minds of their citizens. Such variation might predict the magnitude of the question wording effect (see Van Bavel et al. 2016).

One formal way to identify countries in which mentioning freedom in the survey question might enhance support for free speech the most is to diagnose the centrality of the notion of freedom in the values of a nation. Specifically, countries that attach greater cultural value to the notion of freedom might be most influenced. Fortunately, this can be tested using surveys of representative samples of the residents of nations in which the importance of freedom to people was measured. We did so by using data from the European Social Survey (ESS).

The ESS was conducted in a sufficient number of the countries where we conducted our experiments but not in all of those countries. Therefore, to supplement use of the ESS data, we conducted a second analysis using national anthems to diagnose the cultural values² of all the countries in which our experiments were conducted. This use of national anthems to diagnose the psyche of a nation is in keeping with a scholarly literature that has documented the value of doing so. According to Sondermann (1997), “national anthems are a kind of textual icon [and] the embodiment of the nation” (p. 128; see also Cerulo 1989). Waterman (2019) said that national anthems are “part of the paraphernalia of national packaging” that serve as “a rallying point for expressing personal and group identities.” (p. 2603). Confirming this notion, Rihmer (1997), Voros et al. (2016) and Lester and Gunn (2011) showed that the rates of suicides in various countries can be predicted by the degree to which their national anthems contain sad words and notions of ambivalence, denial, and loss. And experimental evidence has demonstrated that

² In keeping with assumptions made by Schwartz (1992) and Hofstede (1980), we acknowledge that single countries are almost always homes to many subcultures, but it is nonetheless reasonable to assume that residents of a country may share a predominant cultural perspective, which can be characterized in part by which values they tend to prioritize and to which they tend to accord less importance.

listening to one's national anthem evokes associations suggesting influence and internalization (e.g., Gilboa and Bodner 2009).

All of that literature is consistent with social identity theory's central postulate that in the modern world, people are members of multiple social groups based on race, age, gender, and more. And one such social group might, for some people, be the population of the nation where they were raised or where they reside. Identification with social groups is thought to be a way to cultivate positive social identities (Tajfel 1978, 1981; Tajfel and Turner 1979, 1986). One way people might do so is by attaching cultural importance to the values that the nation seems to endorse. In the present context, to the extent that a nation endorses the value of freedom and the residents of the nation internalize that value, those people might be especially motivated by addition of the word "freedom" to Schuman and Presser's (1981) question.

As in America, "freedom" or "free" appear in the national anthems of five other countries in our sample: Canada ("God keep our land glorious and free"), Germany ("Unity and rights and freedom are the basis of good fortune"), the Netherlands ("A prince I am, undaunted, of Orange, ever free"), Taiwan ("Our aim shall be to found a free land"), and Sweden ("Thou free").³ In contrast, the words "freedom" and "free" do not appear in the anthems of five other countries in our sample: France, Denmark, Iceland, Norway, and the United Kingdom.⁴ If residents of the latter countries do indeed attach less importance to freedom than do residents of the former countries, then perhaps the question wording effect observed by Schuman and Presser (1981) will appear more strongly in the former countries than in the latter.

By testing this hypothesis, we conducted a test of the notion that national anthems

³ We focus our discussion of the verses of the national anthems that are routinely sung these days at public events, which are subsets of the verses in the original compositions in many countries.

⁴ Note that "liberty, equality, and fraternity" is the national motto of France; "liberty" is a synonym for "freedom."

embody cultural values of nations. Specifically, we gauged whether the amount of impact of adding the word “freedom” to the question can be predicted by the presence or absence of degree to which residents of each country value freedom. And we assessed the degree to which mention of freedom in national anthems corresponds to endorsement of freedom as a value in national surveys.

We also explored cross-cultural moderation of this question wording effect by cognitive skills. Although Schuman and Presser (1981) did not observe moderation of the strength of question design effects by respondent education (a good proxy for cognitive skills), Narayan and Krosnick (1996) conducted meta-analyses of Schuman and Presser’s (1981) data and documented stronger response effects among less educated respondents (Krosnick 1991). But Narayan and Krosnick (1996) did not look for moderation of the question wording effect involving “freedom,” so we did.

One possibility is that education might not moderate this question wording effect, because the word “freedom” may activate the same concept and cultural resonance among all individuals living in a country that values freedom, regardless of cognitive skills. On the other hand, it is possible that more cognitively skilled individuals are naturally more likely to recognize the relevance of the concept of freedom to the issue at hand, even if the word “freedom” is not explicitly in the survey question. Therefore, inclusion of the word “freedom” may have little impact on highly cognitively skilled respondents and may have more substantial impact on less skilled respondents, by pointing out to them a relevance that they didn’t naturally recognize. We explored these possibilities.

We also examined the role of another factor in influencing the magnitude of the question wording effect across countries: the level of religiosity in a country. If religion is not especially important to people, they are likely to be open to allowing a speech against religion regardless of

whether the notion of freedom is activated by the question or not. But if religion is very important to the residents of a country, they may be initially inclined to oppose a speech against religion. Thus, adding an emphasis on freedom to a question may make these people more open to what they would otherwise oppose.

This paper proceeds as follows. First, we describe the data that were collected to investigate these issues across 11 countries. Then we compare Schuman and Presser's (1981) finding to the results obtained from new data collected in the U.S. to assess replication. Next, we turn to testing whether the effect appeared in the other ten countries and whether its presence was moderated by the value attached to freedom by adults in each country, as measured by: (1) the average importance attached to the value of freedom, as measured by the Schwartz Value Scale (Schwartz 1992) administered in the European Social Survey, and, less formally, (2) the presence of the word "freedom" in the country's national anthem. Then we explore whether education and religiosity of a country moderate the size of the question wording effect. Finally, we test Schuman and Presser's (1981) form-resistant correlation hypothesis, which proposes that the ranking of and spacing between the countries in terms of their attitudes toward allowing the speech may be the same regardless of whether "freedom" was included in or excluded from the question.

Methods

Data

The question wording experiment was implemented using thirteen probability samples of the general population in eleven countries (see Table 1) as part of the Multi-National Study of Questionnaire Design (MSQD, see Silber et al. 2018). Data were collected between 2013 and 2015 via the Internet in the United States (TESS and Gallup), France, Germany (GIP), the Netherlands, Taiwan, Iceland, Norway, and Sweden. A mixed-mode design was employed in

Canada (online and telephone), in Denmark (online, mail, and telephone), in Germany (the GESIS Panel; online and mail), and in the United Kingdom (online and computer-assisted self-completion during a face-to-face interview).⁵ The number of respondents who answered the experimental question varied between 789 and 4,210. In total, 23,370 respondents did so. A detailed description of the study setup, translation procedure, and the sampling strategy in each country is provided by Silber et al. (2018).⁶ Basic methodological information for each study appears in Appendix A, and the question wordings used in each country appear in Appendix B.

The importance attached to freedom and religiosity by members of representative samples of the adults living in each country was gauged using data from the 2014 and 2016 rounds of the European Social Survey (ESS 2014, 2016). Among the countries in which the ESS collected data in 2014 were seven that we studied (Denmark, France, Germany, Netherlands, Norway, Sweden, and the United Kingdom, $N = 13,646$), and among the countries studied by the ESS in 2016 was a slightly different set of seven (France, Germany, Iceland, Netherlands, Norway, Sweden, and the United Kingdom, $N = 12,409$) that correspondent to the countries in our study. We combined the data from the two rounds in order to increase the total number of matched countries to eight, averaging means from pairs of measurements of the same country.

ESS interviews were conducted face-to-face in a probability sample of individuals living

⁵ Japan and Portugal participated in the MSQD but did not administer the experiment described in this paper.

⁶ To optimize translation, the source questionnaire in English was given to the research teams in the various countries, who developed functionally equivalent translations following the TRAPD (Translation, Review, Adjudication, Pretesting, and Documentation) approach (Harkness et al. 2010b). It is useful to note that although the experiment we examined here used Schuman and Presser's wording in English in the United States, Canada (for some respondents), and the United Kingdom, the translations into other languages added the two words "the freedom" only in Germany and Sweden. Translations into other languages involved slightly different constructions. It is tricky business to translate a questionnaire from one language to another (Behr 2017), but it is possible to summarize the way "the freedom" was expressed in the other translations: in Taiwan: "have ... freedom", in the Netherlands: "use his/her freedom-of-expression", in French in Canada: "have the freedom", in Denmark and Norway: "use freedom-of-expression", in Iceland: "give freedom to", in French in France: "get the freedom". Thus, although none of these simply said "the freedom", they all used a form of the word "freedom" and expressed the idea as intended. Therefore, these approximate translations seem unlikely to undermine the value of this experiment.

in private households with at least 1,500 respondents per country. In 2014, the response rate (RR1, AAPOR 2016) varied between 31.4% and 58.6% (Denmark 51.9%, France 50.9%, Germany 31.4%, the Netherlands 58.6%, Norway 53.9%, Sweden 50.1%, the United Kingdom 43.6%), and in 2016, it varied between 30.6% and 53.0% (France 52.4%, Germany 30.6%, Iceland 45.8%, the Netherlands 53.0%, Norway 52.8%, Sweden 43.0%, the United Kingdom 42.8%).⁷

Measures

Surveys we commissioned. Respondents were randomly assigned to answer the question including “the freedom” or omitting those words. Response options were “Yes, allow to speak” (coded 1) and “No, not allowed” (coded 0).

In each country, respondents were asked about the highest level of formal education they had completed and were subsequently classified as having a low, medium, or high level of formal education. The measurement and meaning of education varies across countries (Schneider et al. 2016), so experts from the GESIS methodology center in Mannheim, Germany, determined how to best assign respondents in each country to one of the three education levels (see Appendix C for the measurement of education in the various surveys).

European Social Survey. All respondents of the 2014 ESS and the 2016 ESS answered the question about *freedom*, “Now we will briefly describe some people. Please listen to each description and tell me how much each person is or is not like you.” One description said: “It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.” Response choices were “very much like me” (coded 1), “like me” (coded .8), “somewhat like me” (coded .6), “a little like me” (coded .4), “not like me” (coded .2), and “not

⁷ Full information on fieldwork can be found at <http://www.europeansocialsurvey.org/data/round-index.html>.

like me at all” (coded 0). All respondents also answered the question on *religiosity*, “Regardless of whether you belong to a particular religion, how religious would you say you are?” Responses were provided on an 11-point scale with verbal labels only on the end points: “not at all religious” to “very religious”. Responses were coded to range from 0 to 1. The answers to these two questions were averaged across respondents from each country to yield a score per country.⁸ See Appendix D for the ESS question wordings used in the various countries.

Analyses

Tests of the impact of adding “freedom” to the target question on response distributions were conducted using two ordinary least squares (OLS) regressions in which (1) question form predicted responses to the target question (thus testing what we call the “difference” due to the question wording manipulation), and (2) gender and education were added as predictors (thus testing what we call the “adjusted difference” due to the question wording manipulation), with respondents being the unit of analysis:

$$\hat{y}(\text{response to the target question}) = \beta_1(\text{question form}) + \beta_2(\text{gender}) + \beta_3(\text{education}) \quad (1)$$

In order to compare variation in the size of the question wording effect across countries, the parameters of OLS regression equations were estimated, treating country as the unit of analyses, predicting the size of the question wording effect in each country (i.e., the difference between the percentages of respondents who advocated allowing the speech documented by the two question wordings) using the importance of freedom in the country (measured either with the ESS data or the national anthem codings), the percent of respondents in the country who

⁸ The publicly available ESS dataset for 2014 and 2016 include a design weight (to correct for unequal probability of selection) and a weight that combines the design weight with a post-stratification weight (that corrects for differential non-response). All analyses reported here used the latter weight.

advocated allowing the speech when “freedom” was omitted from the target question (to control for possible ceiling effects), and religiosity of the country, as explained below.

Question Wording Effect(condition 2 – condition 1)

$$\begin{aligned}
 &= \beta_1(\text{cultural importance of freedom}) \\
 &+ \beta_2(\text{support for allowing the speech when "freedom" was omitted}) \\
 &+ \beta_3(\text{religiosity})
 \end{aligned}
 \tag{2}$$

The form-resistant correlations hypothesis was tested by assessing whether the rank ordering of and spacing between countries in terms of the percent of respondents who endorsed allowing the speech were maintained regardless of whether the word “freedom” was included in or excluded from the target question. A Pearson product moment correlation was computed, treating country as the unit of analysis. The closer this correlation is to 1, the less impact question wording had on the ordering and spacing of countries.

Results

Replication in the U.S.

The question order effect documented by Schuman and Presser (1981) appeared in the Gallup data in the U.S. (see Table 2)⁹. 93.7% of respondents endorsed allowing the speech when “the freedom” was omitted, compared to 95.8% when those words were included. This difference of 2.1 percentage points is statistically significant ($t = 2.058, p = .020$). The same effect of question wording was apparent in the TESS data: 82.0% endorsed allowing the speech when “the freedom” was omitted, and 85.8% did so when those words were included. This difference of 3.8 percentage points was marginally significant ($t = 1.640, p = .051$). The effect of

⁹ All p-values for differences in theoretically expected directions are one-tailed. All other p-values are two-tailed. As is apparent in the table, similar results were obtained regardless of whether we examine the differences or the adjusted differences, so in the text, we discuss the differences.

the wording variation was not significantly different in the two samples ($\chi^2(1) = .269, p = .604$). Combining the Gallup and TESS data, the effect of the question wording manipulation in the U.S. is 2.8 percentage points ($t = 2.664, p = .004$).

Generalization Across Other Countries

The effect of adding “freedom” to the question was statistically significant and in the expected direction in four countries: 3.5 percentage points in Canada ($t = 2.269, p = .012$), 4.3 percentage points in Germany in the GESIS panel ($t = 3.715, p < .001$), 8.0 percentage points in Germany in the GIP ($t = 3.529, p < .001$), 4.1 percentage points in the Netherlands ($t = 2.759, p = .003$), and 5.8 percentage points in Taiwan ($t = 2.136, p = .017$). The effects in the two German samples were not significantly different from one another ($\chi^2(1) = 2.461, p = .118$). The within-country replication reinforces the conclusion that the finding is reliable.

The wording manipulation had no significant effect in six countries: -0.9 percentage points in France ($t = -0.331, p = .740$), -0.2 percentage points in the UK ($t = -0.141, p = .888$), -0.1 percentage points in Iceland ($t = -0.119, p = .905$), 0.5 percentage points in Denmark ($t = 0.370, p = .356$), 1.2 percentage points in Sweden ($t = 1.100, p = .131$), and 1.9 percentage points in Norway ($t = 1.107, p = .134$).

Tests of Moderation Using a Direct Measure of the Importance of Freedom

To test the moderation hypothesis formally, we estimated the parameters of Equation (2) above using the average importance attached to freedom by the adult residents of the country, as gauged by the Schwartz (1992) measure in the ESS. This reduced the number of analyzed countries from 11 to 8. The effect of the value of freedom on the effect size of the question wording manipulation was positive and statistically significant ($b = 41.931, CI\ 95\% = 15.349, 68.512, p = .014$; see Model 1 in Table 3), meaning that the wording effect was greater in countries that attached more value to freedom. Thus, the question wording effect was moderated

by these values as expected.

In order to conduct a second test with all eleven of the countries, the presence of freedom in the countries' national anthems was substituted for the ESS measure of values. Across the eight countries in which the ESS measured values, the point-biserial correlation between those measurements and the presence/absence of "freedom" in the countries' national anthems was positive and substantial, as expected: .584. Thus, there is considerable overlap between these measures, but they are not identical.

As expected, the question wording effect was positive and statistically significant when combining the countries in which the word "freedom" is in the national anthem (3.8 percentage points, $\chi^2 (1) = 45.955, p < .001$), and the effect was not statistically significant in the other countries where freedom was not in the national anthem (0.2 percentage points, $\chi^2 (1) = 0.115, p = .734$). The question wording effect differed statistically significantly between these two groups of countries by 3.6 percentage points ($\chi^2 (1) = 15.264, p < .001$). The question wording effect was homogeneous (meaning that the effect did not differ in size across these countries) among the countries whose national anthems included freedom ($\chi^2 (5) = 0.639, p = .986$) and among the countries whose national anthems did not include it ($\chi^2 (4) = 1.385, p = .847$).

In Equation (2), when the presence/absence of "freedom" in the country's national anthem replaced the ESS measure of the value of freedom in the countries, this predictor had the expected, positive statistically significant effect on the size of the question wording effect ($b = 3.393, CI 95\% = 1.997, 4.788, p = .001$; see Model 2 in Table 3). This result reinforces the conclusion that the question wording effect was greater in countries that attached more value to freedom.

Ceiling Effects and Religiosity

The proportion of people who advocated allowing the speech was very high in all

countries when the question wording omitted the word “freedom”: 79.3% to 94.2%. These numbers are notably larger than the one observed in Schuman and Presser’s (1981) study (65.8%). These high percentages limited the amount that the observed percentage could increase due to the addition of the word “freedom” to the question. And because the starting point percentage was different in different countries, variation in the suppression of the wording effect due to a ceiling may have distorted differences in the size of the wording effect across countries (see, e.g., Wang et al. 2008). Consistent with this reasoning, the correlation between the percent of people advocating allowing the speech when measured with the question omitting the word “freedom” and the size of the effect of adding the word “freedom” is $-.404$.

As expected, the higher the baseline level of support for allowing the speech, the smaller the increase in the percent of respondents advocating allowing the speech due to addition of the word “freedom” to the question (Model 1: $b = -.231$, 95% $CI = -.459, -.003$, $p = .049$; Model 2: $b = -.154$, 95% $CI = -.293, -.014$, $p = .038$). Also as expected, the more religious a country was, the less willing its residents were to allow a speech against religion when asked the question that did not include the word “freedom”: $r = -.515$ when treating country as the unit of analysis.¹⁰ However, the religiosity of a country did not significantly predict the size of the question wording effect ($b = -15.407$, 95% $CI = -36.385, 5.570$, $p = .096$; see Model 1 in Table 3)

Moderation by Education

Combining the data from the countries in which the question wording effect appeared significantly, that wording effect was the same in the low and medium education groups

¹⁰ This correlation is based on the eight countries for which country-level variables on freedom were available in the ESS.

($b = -.003$, $CI\ 95\% = -.035, .029$, $p = .867$).¹¹ However, the question wording effect was weaker among the high education respondents than among the medium and low education respondents combined ($b = -.027$, $95\%\ CI = -.048, -.006$, $p = .019$).¹² This is consistent with the conclusion that the highly educated respondents recognized the relevance of freedom of speech to the question even when the word “freedom” was not explicitly included.

Form-resistant Correlation Hypothesis

Schuman and Presser’s (1981) “form-resistant correlation” hypothesis proposes that substantive responses to a question will correlate 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 will correlate 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 of and spacing between countries in terms of their responses to the speech question were maintained regardless of whether the word “freedom” was included or excluded. Treating country as the unit of analysis, the Pearson product moment correlation between the numbers in the first two columns of Table 4 is .901. This extremely strong correlation sustains the form-resistant correlation hypothesis and attests to the robustness of the ordering and spacing of the countries in terms of endorsement of free speech opportunities.

¹¹ This difference was tested by estimating the parameters of an OLS regression equation predicting responses to the question about the speech using three predictors: a dummy variable (coded 0 or 1) indicating which question wording the respondent received, a dummy variable coded 0 for low education respondents and 1 for medium education respondents, and the interaction of those two variables. High education respondents were not included in this regression.

¹² This difference was tested by estimating the parameters of an OLS regression equation predicting responses to the question about the speech using three predictors: a dummy variable (coded 0 or 1) indicating which question wording the respondent received, a dummy variable coded 0 for low and medium education respondents and 1 for high education respondents, and the interaction of those two variables.

Discussion

Summary of Findings

This study documented that a question wording effect first reported 40 years ago with a representative sample of Americans was replicated recently in two representative samples of residents of the same country. Furthermore, the effect generalized to nations that appear to place great cultural value on the notion of freedom, where the question wording effect was statistically significant and in the expected direction. In contrast, the effect was not apparent in countries that seem not to place the same explicit emphasis on freedom in their national culture.

In the countries where the question wording effect appeared, it was most pronounced among people with moderate and low levels of education, consistent with the idea that highly educated individuals are most likely to recognize the relevance of freedom to the issue at hand, even if the word “freedom” is not explicitly mentioned in the survey question. The extremely strong evidence supporting the form-resistant correlation hypothesis suggests that the ranking and spacing of countries in terms of endorsement of free speech is very robust and equally apparent regardless of question wording.

Contribution to the Question Design Literature

This evidence makes a useful contribution to the huge array of studies of question wording in the survey methods literature (see Krosnick and Fabrigar forthcoming). Some past experiments have demonstrated no effects of changing wording. For example, the distributions of attitudes toward abortion remained identical when that word was replaced with the less inflammatory “end pregnancy” (Schuman and Presser 1981). Similarly, Schaeffer (1982) found that people answered a question about values for child-rearing similarly regardless of whether the question used the gendered word “he” or the neutral phrase “a child”.

In contrast, other studies have found that words that survey designers might think are

synonyms were not so in the minds of survey respondents. For example, Rasinski (1989) found that more people endorsed government spending on “dealing with drug addiction” than on “drug rehabilitation”. More respondents endorsed “military training” than endorsed “training for war” (Rugg and Cantril 1944). And seemingly innocuous changes in the wording of a question about President Barack Obama’s birthplace yielded strikingly different results (Krosnick et al. 2014). Thus, sometimes, a seemingly simple question wording shift matters, and sometimes it does not. The present study illustrates how and why and where a particular question wording shift is consequential and suggests that the process varies depending not only on the meanings of the words involved but also on the cultural context in which the question is asked.

How should investigators word a question intended to measure attitudes toward free speech? Should the word “freedom” be included, or should it be omitted? If a researcher’s goal is to measure opinions with maximum accuracy and minimum bias, this decision seems consequential, because different results are obtained depending on which approach one takes. Fortunately, the results are not very different, and the ranking and spacing of countries one obtains are nearly identical regardless of which question wording is employed. Still, a researcher setting out to measure opinions optimally can legitimately ask for guidance about which wording to use.

We are inclined to believe that both wordings are legitimate but have different meanings. One wording allows a researcher to gauge opinions when the notion of freedom has been cognitively activated by the wording alteration, and the other wording allows for assessment of opinions when the notion of freedom has not been activated in this way. In some countries, these two opinions appear to be the same, and in other countries, they appear to be different. Thus, results obtained with both question wordings have value for understanding opinions in this domain. This is yet another demonstration that, at least in some contexts, question wording

matters (e.g., Newport 2004), so researchers should always be sure to interpret and describe results using the exact wording used in a question, rather than assuming that the wording can be paraphrased without causing damage (Krosnick 1989)

Implications for the Literature on Replication

These findings cast an interesting light on recent attempts to “replicate” social science findings cross-nationally (e.g., Ebersole et al. 2016; Klein et al. 2014). In his early study of question wording effects, Rugg (1941) found that more American respondents said that speeches against democracy should not be allowed than said they should be forbidden. The fact that the same effect was observed in 36 replications conducted in ten countries and seven languages (Klein et al. 2014) increases confidence in the generalizability of the effect.

In that light, it might at first seem troubling that the effect of “freedom” did not appear in all of the countries we examined. The conventional approach to such a pattern of results would be to view them as indicating the original result was fragile and perhaps not even real. But the findings reported here show that the presence and strength of the question wording effect depends quite sensibly on the salience and endorsement of freedom in each country and on *a priori* attitudes toward allowing speeches against religion. Consequently, the variability reported here should probably be viewed as indicating limits on the generalizability of a question wording effect rather than uncertainty about its reliability. Indeed, in each pair of datasets collected in the same country (the U.S. and Germany), similar question wording effects were observed, evidence appropriately viewed as documenting the replicability of the finding.

This conclusion about meaningful variation across countries in the question wording effect resonates with similar, meaningful cross-country variation in question order effects documented by Stark et al. (2018). In their work, the size of question order effects varied as expected with the extent to which necessary conditions for the question order effect to appear

were met in the country. Therefore, the evidence reported in this paper illustrates why before concluding that many failed attempts to replicate a finding calls into question the reality of that finding (see, e.g., Open Science Collaboration et al. 2015), it is useful to explore whether instances of non-replication indicate lack of generalization rather than lack of robustness of the original demonstration (see also Van Bavel et al. 2016; Stroebe and Strack 2014).

Contribution to the Literature on National Anthems

The findings reported here contribute usefully to the growing literature on national anthems. In recent decades, scholars have found national anthems to be useful indicators with which to understand the psyche of nations (Cerulo 1989; Gilboa and Bodner 2009; Lester and Gunn 2011; Rihmer 1997; Sondermann 1997; Voros et al. 2016; Waterman 2019). The present evidence reinforces this notion. Specifically, the presence of “freedom” in a country’s anthem correlated positively with the value to which residents of the country attach to freedom, as measured by survey data from representative samples. And experimental evidence documented that mentioning freedom has especially potent impact on countries whose national anthems include that word. All this constitutes strong convergence on the notion that the study of national anthems is worthwhile and may yield useful insights in the future.

Contribution to the Literature on Cultural Values

In recent decades, a huge literature has emerged studying cultural values as a tool for understanding differences between nations. An important consequence of that literature has been the inclusion of cultural values measures in the European Social Survey, thereby giving countless investigators helpful empirical tools for the continued study of this phenomenon. This collection of quantitative data to characterize the cultural values of countries has yielded many important publications (e.g., Besley 2008; Bilsky et al. 2011; Datler et al. 2013; Davidov 2008; Davidov et al. 2008; Kuntz et al. 2015; Piurko et al. 2011; Saris et al. 2013; Schwartz and Rubel-Lifschitz

2009). The present study makes use of those data and yielded findings that contribute to that literature by validating the survey-based measure of the importance that national populations attach to “freedom” and illustrating a new instance in which cultural values are helpful for explaining social phenomena.

Caveats

It is interesting to consider the present evidence from the perspective of linguist Kenneth Pike (1967), who distinguished between emic and etic approaches to the study of cultures. Summarized over-simplistically, an emic approach to studying social behavior embeds an investigation within one culture at a time and digs deeply to identify local richness. In contrast, an etic approach seeks to keep a bit more distance from each culture studied and focuses on comparing cultures to one another while presuming some universals shared by all cultures. The present investigation can be viewed as etic in spirit, meaning that we presumed that the notion of freedom has been understood by people consistently over the countries we studied and over the decades between the 1970s (when Schuman and Presser collected their data) and recent years. The variation across countries in the importance seemingly attached to the notion of freedom is not inconsistent with the presumption that people in those places and times generally understood the concept similarly. And in light of that assumption, the results reported here seem sensible. But it is important to acknowledge that they are based on the etic perspective.

Another way to think about the analysis done here is in terms of Harkness et al.’s (2010a) model of the relations among concepts, constructs, indicators, and questions in cross-national surveys (see p. 42). Schuman and Presser’s (1981) experiment varied the presence of the word “freedom” in the question, meaning that the study examined the impact of question variation on responses. But it is unclear whether the cross-national differences highlighted by the current evidence are due to differences across countries in the importance ascribed to a single construct

or concept (that is equivalently present in all countries) or to different constructs or concepts being measured in different countries. That is, perhaps for Americans, the notion of freedom refers to the rights of individuals within the country to live their lives with minimal government restrictions, whereas for people in Taiwan, “freedom” refers to their nation’s ability to carry out its business with minimal interference from neighboring countries. The findings reported here set the stage for future investigation of such issues.

Limitations

The use of probability samples in the present study represents a methodological improvement over prior studies that instead examined haphazard, convenience samples in various countries (e.g., Ebersole et al. 2016; Klein et al. 2014). However, the particular set of countries in which data were collected for the present study is not a random subset of countries, so there is reason to hesitate before generalizing these results to other countries. Exploring the replicability of the findings reported here in other countries would be well worth doing.

Another limitation of the present study is that 10 of the 11 countries studied are “Western”, meaning that it will be interesting in the future to explore these same issues in other nations around the world. Although one might wonder whether there is sufficient variation across the countries we studied to allow a study of cultural influence, we saw sufficient variability across national surveys (reinforced by the national anthems) to allow successful prediction of the size of the question wording effects. Thus, it appears that variation was sufficient for the present purposes.

Code and Data Availability

All code and data to reproduce the results of this paper and supplementary tables, is available at <https://osf.io/mw6ks/>.

References

- AAPOR (2016), "Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys," Available at https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.
- Baker, M. (2016), "Biotech giant publishes failures to confirm high-profile science," *Nature*, 530, 141. <https://doi.org/10.1038/nature.2016.19269>.
- Bargh, J. A., Lee-Chai, A., Barndollar, K., Gollwitzer, P. M., and Trötschel, R. (2001), "The automated will: Nonconscious activation and pursuit of behavioral goals," *Journal of Personality and Social Psychology*, 81, 1014. <https://doi.org/10.1037/0022-3514.81.6.1014>.
- Behr, D. (2017), "Assessing the use of back translation: the shortcomings of back translation as a quality testing method," *International Journal of Social Research Methodology*, 20, 573–584. <https://doi.org/10.1080/13645579.2016.1252188>.
- Besley, J. C. (2008), "Media use and human values," *Journalism and Mass Communication Quarterly*, 85, 311–330. <https://doi.org/10.1177/107769900808500206>.
- Bilsky, W., Janik, M., and Schwartz, S. H. (2011), "The structural organization of human values—evidence from three rounds of the European Social Survey (ESS)," *Journal of Cross-Cultural Psychology*. <https://doi.org/10.1177/0022022110362757>.
- Blom, A. G., Gathmann, C., and Krieger, U. (2015), "Setting Up an Online Panel Representative of the General Population: The German Internet Panel," *Field Methods*, 27, 391–408. <https://doi.org/10.1177/1525822X15574494>.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., and Wu, H. (2016), "Evaluating replicability of laboratory experiments in economics," *Science*, 351, 1433–1436.

<https://doi.org/10.1126/science.aaf0918>.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E. J., and Wu, H. (2018), “Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015,” *Nature Human Behaviour*, 2, 637. <https://doi.org/10.1038/s41562-018-0399-z>.

Carey, B. (2015), “Many psychology findings not as strong as claimed, study says,” *New York Times*, Available at https://www.nytimes.com/2015/08/28/science/many-social-science-findings-not-as-strong-as-claimed-study-says.html?_r=0.

Cerulo, K. A. (1989), “Sociopolitical control and the structure of national symbols: An empirical analysis of national anthems,” *Social Forces*, 68, 76–99. <https://doi.org/10.1093/sf/68.1.76>.

Cornesse, C., and Blom, A. G. (2020), “Response quality in nonprobability and probability-based online panels,” *Sociological Methods & Research*.
<https://doi.org/10.1177/0049124120914940>

Cornesse, C., Blom, A. G., Dutwin, D., Krosnick, J. A., de Leeuw, E., Legleye, S., Pasek, J., Pennay, D., Phillips, B., Sakshaug, J. W., Struminskaya, B., and Wenz, A. (2020), “A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research.” *Journal of Survey Statistics and Methodology*, 8(1), 4–36.
<https://doi.org/10.1093/jssam/smz041>

Datler, G., Jagodzinski, W., and Schmidt, P. (2013), “Two theories on the test bench: Internal and external validity of the theories of Ronald Inglehart and Shalom Schwartz,” *Social Science Research*, 42, 906–925. <https://doi.org/10.1016/j.ssresearch.2012.12.009>.

Davidov, E. (2008), “A cross-country and cross-time comparison of the human values

measurements with the second round of the European social survey,” *Survey Research Methods*, 2, 33–46. <https://doi.org/10.18148/srm/2008.v2i1.365>.

Davidov, E., Schmidt, P., and Schwartz, S. H. (2008), “Bringing values back in: The adequacy of the European social survey to measure values in 20 countries,” *Public Opinion Quarterly*, 72, 420–445. <https://doi.org/10.1093/poq/nfn035>.

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., Davis, W. E., Devos, T., Fletcher, M. M., German, K., Grahe, J. E., Hermann, A. D., Hicks, J. A., Honeycutt, N., Humphrey, B., Janus, M., Johnson, D. J., Joy-Gaba, J. A., Juzeler, H., Keres, A., Kinney, D., Kirshenbaum, J., Klein, R. A., Lucas, R. E., Lustgraaf, C. J. N., Martin, D., Menon, M., Metzger, M., Moloney, J. M., Morse, P. J., Prislín, R., Razza, T., Re, D. E., Rule, N. O., Sacco, D. F., Sauerberger, K., Shrider, E., Shultz, M., Siemen, C., Sobocko, K., Weylin Sternglanz, R., Summerville, A., Tskhay, K. O., van Allen, Z., Vaughn, L. A., Walker, R. J., Weinberg, A., Wilson, J. P., Wirth, J. H., Wortman, J., and Nosek, B. A. (2016), “Many Labs 3: Evaluating participant pool quality across the academic semester via replication,” *Journal of Experimental Social Psychology*, 67, 68–82. <https://doi.org/10.1016/j.jesp.2015.10.012>.

ESS (2014), *ESS Round 7: European Social Survey Round 7 Data file edition 2.1*, NSD - Norwegian Centre for Research Data, Norway.

ESS (2016), *ESS Round 8: European Social Survey Round 8 Data file edition 1.0*, NSD - Norwegian Centre for Research Data, Norway.

Gilboa, A., and Bodner, E. (2009), “What are your thoughts when the national anthem is playing? An empirical exploration,” *Psychology of Music*, 37, 459–484.

<https://doi.org/10.1177/0305735608097249>.

Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., and Villar, A. (2010a), "Designing questionnaires for multipopulation research," in *Survey Methods in Multicultural, Multinational, and Multiregional Contexts*, eds. J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, and T. W. Smith, Hoboken, NJ: Wiley, pp. 115–140.

Harkness, J. A., Villar, A., and Edwards, B. (2010b), "Translation, Adaptation, and Design," in *Survey Methods in Multicultural, Multinational, and Multiregional Contexts*, eds. J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. E. Lyberg, P. P. Mohler, B.-E. Pennell, and T. W. Smith, Hoboken, NJ: Wiley, pp. 115–140.

<https://doi.org/10.1002/9780470609927.ch7>.

Hofstede, G. (1980). *Culture's consequences*, Beverly Hills, CA: Sage.

Hyman, H. H., and Sheatsley, P. B. (1950), "The current status of American public opinion," in *The teaching of contemporary affairs*, ed. J. C. Payne, New York, NY: National Education Association, pp. 11–34.

Ioannidis, J. P. A. (2005), "Why most published research findings are false," *PLoS medicine*, 2, e124. <https://doi.org/10.1371/journal.pmed.0020124>.

John, L. K., Loewenstein, G., and Prelec, D. (2012), "Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling," *Psychological Science*, 23, 524–532. <https://doi.org/10.1177/0956797611430953>.

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., Hasselman, F., Hicks, J. A., Hovermale, J. F., Hunt, S. J., Huntsinger, J. R., Ijzerman, H.,

- John, M. S., Joy-Gaba, J. A., Kappes, H. B., Krueger, L. E., Kurtz, J., Levitan, C. A., Mallett, R. K., Morris, W. L., Nelson, A. J., Nier, J. A., Packard, G., Pilati, R., Rutchick, A. M., Schmidt, K., Skorinko, J. L., Smith, R., Steiner, T. G., Storbeck, J., Van Swol, L. M., Thompson, D., Van 'T Veer, A. E., Vaughn, L. A., Vranka, M., Wichman, A. L., Woodzicka, J. A., and Nosek, B. A. (2014), "Investigating variation in replicability: A 'many labs' replication project," *Social Psychology*, 45, 142–152.
<https://doi.org/10.1027/1864-9335/a000178>.
- Krosnick, J. A. (1989), "A Review: Question Wording and Reports of Survey Results: The Case of Louis Harris and Associates and Aetna Life and Casualty," *Public Opinion Quarterly*, 53, 107–113. <https://doi.org/10.1086/269144>.
- Krosnick, J. A. (1991), "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys," *Applied Cognitive Psychology*, 5, 213–236.
<https://doi.org/https://doi.org/10.1002/acp.2350050305>.
- Krosnick, J. A., and Fabrigar, L. R. (forthcoming). *The handbook of questionnaire design*, New York, NY: Oxford University Press.
- Krosnick, J. A., Malhotra, N., and Mittal, U. (2014), "Public misunderstanding of political facts: How question wording affected estimates of partisan differences in birtherism," *Public Opinion Quarterly*, 78, 147–165. <https://doi.org/10.1093/poq/nft080>.
- Kuntz, A., Davidov, E., Schwartz, S. H., and Schmidt, P. (2015), "Human values, legal regulation, and approval of homosexuality in Europe: A cross-country comparison," *European Journal of Social Psychology*, 45, 120–134. <https://doi.org/10.1002/ejsp.2068>.
- Lester, D., and Gunn, J. F. (2011), "National anthems and suicide rates," *Psychological Reports*, 108, 43–44. <https://doi.org/10.2466/12.PR0.108.1.43-44>.
- Martinsson, J., Andreasson, M., Markstedt, E., and Riedel, K. (2014), *Technical Report Citizen*

Panel 9 - 2014, Gothenburg: University of Gothenburg, LORE.

Narayan, S., and Krosnick, J. A. (1996), "Education moderates some response effects in attitude measurement," *Public Opinion Quarterly*, 60, 58–88.

<https://doi.org/https://doi.org/10.1086/297739>.

Newport, F. (2004), *Polling matters: Why leaders must listen to the wisdom of the people*, New York, NY: Time Warner Book Group.

Open Science Collaboration, O. S., Hempel, C., Hempel, C., Oppenheim, P., Meehl, P. E., Platt, J. R., Nosek, B. A., Lakens, D., Schmidt, S., Ioannidis, J. P. A., Begley, C. G., Ellis, L. M., Prinz, F., Schlange, T., Asadullah, K., McNutt, M., Pashler, H., Wagenmakers, E.-J., Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., Munafò, M. R., Fanelli, D., Greenwald, A. G., Howard, G. S., Lau, M. Y., Maxwell, S. E., Venter, A., Lundy, R., Sweeny, R. M., Ioannidis, J. P. A., Munafò, M. R., Fusar-Poli, P., Nosek, B. A., David, S. P., John, L. K., Loewenstein, G., Prelec, D., Nosek, B. A., Spies, J. R., Motyl, M., Rosenthal, R., Rozin, P., Simmons, J. P., Nelson, L. D., Simonsohn, U., Fisher, R. A., Viechtbauer, W., Braver, S. L., Thoemmes, F. J., Rosenthal, R., Simonsohn, U., Lakens, D., Klein, R., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., Hasselman, F., Hicks, J. A., Hovermale, J. F., Hunt, S. J., Huntsinger, J. R., IJzerman, H., John, M.-S., Joy-Gaba, J. A., Kappes, H. B., Krueger, L. E., Kurtz, J., Levitan, C. A., Mallett, R. K., Morris, W. L., Nelson, A. J., Nier, J. A., Packard, G., Pilati, R., Rutchick, A. M., Schmidt, K., Skorinko, J. L., Smith, R., Steiner, T. G., Storbeck, J., Swol, L. M. Van, Thompson, D., Veer, A. E. van 't, Vaughn, L. A., Vranka, M., Wichman, A. L., Woodzicka, J. A., Nosek, B. A., Kappes, H. B., Krueger, L. E., Kurtz, J., Levitan, C. A., Mallett, R. K.,

Morris, W. L., Nelson, A. J., Nier, J. A., Packard, G., Pilati, R., Rutchick, A. M., Schmidt, K., Skorinko, J. L., Smith, R., Steiner, T. G., Storbeck, J., Swol, L. M. Van, Thompson, D., Veer, A. E. van 't, Vaughn, L. A., Vranka, M., Wichman, A. L., Woodzicka, J. A., Nosek, B. A., Cohen, J., Sterling, T. D., Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., Nosek, B. A., Hartshorne, J. K., Schachner, A., Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T. A., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E. L., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., Yarkoni, T., Rosenthal, R., Fode, K. L., Bressan, P., Stranieri, D., Albarracín, D., Handley, I. M., Noguchi, K., McCulloch, K. C., Li, H., Leeper, J., Brown, R. D., Earl, A., Hart, W. P., and Cumming, G. (2015), "Estimating the reproducibility of psychological science.," *Science (New York, N.Y.)*, 349, aac4716. <https://doi.org/10.1126/science.aac4716>.

Pashler, H., and Wagenmakers, E. J. (2012), "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?," *Perspectives on Psychological Science*, 7, 528–530. <https://doi.org/10.1177/1745691612465253>.

Pike, K. L. (1967), *Language in Relation to a Unified Theory of the Structure of Human Behavior*, The Hague, Netherlands: Mouton. <https://doi.org/10.1515/9783111657158>.

Piurko, Y., Schwartz, S. H., and Davidov, E. (2011), "Basic Personal Values and the Meaning of Left-Right Political Orientations in 20 Countries," *Political Psychology*, 32, 537–561. <https://doi.org/10.1111/j.1467-9221.2011.00828.x>.

Rasinski, K. A. (1989), "The Effect of Question Wording on Public Support for Government

- Spending,” *Public Opinion Quarterly*, 53, 388–394. <https://doi.org/10.1086/269158>.
- Rihmer, Z. (1997), “Studies of Suicide and Suicidal Behavior in Hungary,” in *Basic and Clinical Science of Mental and Addictive Disorders*, eds. L. L. Judd, B. Saletu, and V. Filip, Basel: Karger, pp. 171–174. <https://doi.org/10.1159/000059513>.
- Rugg, D. (1941), “Experiments in Wording Questions: II,” *Public Opinion Quarterly*, 5, 91–92. <https://doi.org/10.1086/265467>.
- Rugg, D., and Cantril, H. (1944), “The wording of questions,” in *Gauging public opinion*, ed. H. Cantri, New Jersey: Princeton University Press, pp. 23–50.
- Saris, W. E., and Gallhofer, I. N. (2007), *Design, evaluation, and analysis of questionnaires for survey research*, Hoboken, NJ: Wiley.
- Saris, W. E., Knoppen, D., and Schwartz, S. H. (2013), “Operationalizing the theory of human values: Balancing homogeneity of reflective items and theoretical coverage,” *Survey Research Methods*, 7, 29–44. <https://doi.org/10.18148/srm/2013.v7i1.5040>.
- Schaeffer, N. C. (1982), “A General Social Survey Experiment in Generic Words,” *Public Opinion Quarterly*, 46, 572–581. <https://doi.org/10.1086/268753>.
- Schneider, S., Joye, D., and Wolf, C. (2016), “When translation is not enough: background variables in comparative surveys,” in *The SAGE handbook of survey methodology*, eds. C. Wolf, D. Joye, T. W. Smith, and Y. -c. Fu, London: Sage, pp. 288–307.
- Schooler, J. (2011), “Unpublished results hide the decline effect,” *Nature*, 470(7335), 437–438. <https://doi.org/10.1038/470437a>.
- Schuman, H., and Presser, S. (1981), *Questions and answers in attitude surveys: Experiments on question form, wording, and context*, San Diego, CA: Academic Press.
- Schwartz, S. H. (1992), “Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries,” *Advances in Experimental Social Psychology*, 25, 1–65.

[https://doi.org/10.1016/S0065-2601\(08\)60281-6](https://doi.org/10.1016/S0065-2601(08)60281-6).

Schwartz, S. H., and Rubel-Lifschitz, T. (2009), “Cross-National Variation in the Size of Sex Differences in Values: Effects of Gender Equality,” *Journal of Personality and Social Psychology*, 97, 171–185. <https://doi.org/10.1037/a0015546>.

Silber, H., Stark, T. H., Bloom, A. G., and Krosnick, J. A. (2018), “Multi-national Study of Questionnaire Design,” in *Advances in Comparative Survey Methods: Multicultural, Multinational and Multiregional (3MC) Contexts*, eds. T. P. Johnson, B.-E. Pennell, I. Stoop, and B. Dorer, Hoboken, NJ: Wiley, pp. 161–179.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011), “False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant,” *Psychological Science*, 22, 1359–1366.

<https://doi.org/10.1177/0956797611417632>.

Sondermann, K. (1997), “Reading politically: national anthems as textual icons,” in *Interpreting the Political: New Methodologies*, eds. T. Carver and M. Hyvärinen, London: Routledge, pp. 128–142.

Stark, T. H., 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., and Yu, R. (2018), “Generalization of Classic Question Order Effects Across Cultures,” *Sociological Methods & Research*, in press.

<https://doi.org/10.1177/0049124117747304>.

Stroebe, W., and Strack, F. (2014), “The Alleged Crisis and the Illusion of Exact Replication,” *Perspectives on Psychological Science*, 9, 59–71.

<https://doi.org/10.1177/1745691613514450>.

Tajfel, H. (1978), *Differentiation between social groups: Studies in the social psychology of*

intergroup relations, London: Academic Press.

Tajfel, H. (1981), *Human groups and social categories: Studies in social psychology*, Cambridge: Cambridge University Press.

Tajfel, H., and Turner, J. C. (1979), "An integrative theory of intergroup conflict," in *The social psychology of intergroup relations*, eds. W. G. Austin and S. Worchel, Monterey, CA: Brooks/Cole, pp. 33–48.

Tajfel, H., and Turner, J. C. (1986), "The social identity theory of intergroup behaviour," in *Psychology of intergroup relations*, eds. S. Worchel and W. G. Austin, Chicago: Nelson-Hall, pp. 7–24.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., and Reinero, D. A. (2016), "Contextual sensitivity in scientific reproducibility," *Proceedings of the National Academy of Sciences*, 113, 6454–6459. <https://doi.org/10.1073/pnas.1521897113>.

Voros, V., Osvath, P., Vincze, O., Pusztay, K., Fekete, S., and Rihmer, Z. (2016), "Word use and content analysis of the first verses of six national anthems: A transcultural aspect of suicidal behaviour," *Psychiatria Danubina*, 28, 82–85.

Wang, L., Zhang, Z., McArdle, J. J., and Salthouse, T. A. (2008), "Investigating ceiling effects in longitudinal data analysis," *Multivariate Behavioral Research*, 43, 476–496. <https://doi.org/10.1080/00273170802285941>.

Waterman, S. (2019), "National Anthems and National Symbolism: Singing the Nation," in *Handbook of the Changing World Language Map*, eds. S. D. Brunn and R. Kehrein, Cham, Switzerland: Springer, pp. 2603–2618. https://doi.org/10.1007/978-3-319-73400-2_102-1.

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., and Albarracín, D. (2016), "From primed concepts to action: A meta-analysis of the behavioral effects of incidentally presented words," *Psychological Bulletin*, 142, 472. <https://doi.org/10.1037/bul0000030>.

Tables

Table 1

Percent of Respondents Who Did Not Answer the Target Question and Number of Respondents with Non-Missing Values

Country	% Who Did Not Answer the Question	Total Number of Respondents with Non-Missing Values	Survey Mode ^a
U.S. (Gallup)	1.3	1,986	O
U.S. (TESS)	0.8	1,021	O
Germany (GIP)	8.7	1,038	O
Germany (GESIS)	2.0	4,210	O, M
Taiwan	0.1	789	O
Netherlands	0.7	2,242	O
Canada	0.0	1,317	O
Norway	2.9	1,597	O
Sweden	1.9	1,737	O
Denmark	0.5	1,319	O, T, M
Iceland	2.9	3,051	O
United Kingdom	1.1	2,236	O, CAPI
France	1.0	827	O
Total	1.9	23,370	

^aO = Online, T = Telephone, M= Mail, CAPI = Computer-Assisted Personal Interview

Table 2
*Difference Between the Percent of Respondents Who Said that the Speech Should be Allowed
 When the Word “Freedom” Was Omitted or Included*

Sample	Percent of Respondents Who Said the Speech Should be Allowed				Difference ^a	Adjusted Difference ^b	<i>N</i> _{Adj.}
	Freedom Omitted	<i>N</i> ₁	Freedom Included	<i>N</i> ₂			
Schuman & Presser (1981)	65.8%	503	71.7%	487	-5.9*		
U.S. (Gallup)	93.7	989	95.8	997	-2.1*	-1.8*	1,986
U.S. (TESS)	82.0	522	85.8	499	-3.8 ⁺	-4.3*	1,021
Germany (GIP)	80.2	515	88.1	523	-8.0***	-8.2***	955
Germany (GESIS)	80.9	2,066	85.2	2,144	-4.3***	-4.1***	4,170
Taiwan	79.3	392	85.1	397	-5.8*	-5.5*	789
Netherlands	83.4	1,094	87.5	1,148	-4.1**	-4.4**	2,234
Canada	89.7	670	93.2	647	-3.5*	-3.8**	1,311
Sweden	94.2	876	95.4	861	-1.2	-1.2	1,735
Norway	84.9	790	86.9	807	-1.9	-2.6	1,016
Denmark	92.5	663	93.0	656	-0.5	-0.4	1,319
Iceland	90.5	1,520	90.3	1,531	0.1	0.1	2,919
United Kingdom	86.5	1,097	86.3	1,039	0.2	0.9	2,038
France	79.7	409	78.7	418	0.9	0.9	797

^aThe significance tests are based on OLS regressions with responses to the target question using the question wording dummy variable.

^bThe significance tests and the adjusted differences are based on OLS regressions predicting responses to the target question using the question wording dummy variable, education, and gender.

*** $p < .001$ ** $p < .01$ * $p < .05$ + $p < .10$

Table 3
OLS Regressions Predicting the Size of the Question Wording Effect

Predictor	Model 1			Model 2		
	<i>b</i>	95% <i>CI</i>	<i>p</i>	<i>b</i>	95% <i>CI</i>	<i>p</i>
Value attached to freedom (ESS)	41.931	15.349; 68.512	.014			
Freedom in the national anthem				3.393	1.997; 4.788	.001
Religiosity	-15.407	-.36.385; 5.570	.096			
Percent of respondents who said the speech should be allowed when freedom was omitted from the question	-.231	-.459; -.003	.049	-.154	-.294; -.014	.038
<i>R</i> ²	.794			.765		
Number of countries ^{a)}	8			11		

^{a)}Country was treated as the unit of analysis when computing the OLS regressions (N = 8 in Model 1 and N = 11 in Model 2).

Table 4
Comparing the Rank Ordering and Spacing of the Countries in Terms of the Percent of Respondents Who Said that the Speech Should be Allowed When the Word “Freedom” Was Omitted or Included

Country	Percent of Respondents Who Said the Speech Should be Allowed		N
	Freedom Omitted	Freedom Included	
U.S.	89.7%	92.4%	3,007
Germany	80.8	85.8	5,248
Taiwan	79.3	85.1	789
Netherlands	83.4	87.5	2,242
Canada	89.7	93.2	1,317
Sweden	94.2	95.4	1,737
Norway	84.9	86.9	1,597
Denmark	92.5	93.0	1,319
Iceland	90.5	90.3	3,051
United Kingdom	86.5	86.3	2,236
France	79.7	78.7	827
r^a	.901		

^aCountry was treated as the unit of analysis when computing the Pearson product moment correlation (N = 11).

**Online Appendix of the Manuscript “Replicability of a Classic Question Wording Effect
and Generalization Across Cultures”**

Appendix A. Study Description for Each Data Collection

USA (Gallup Panel)

Survey Sponsor (who paid for the survey). The data collection was funded by the Gallup organization.

Data Collection Organization (who collected the data). The data was collected by the Gallup Organization.

Local Survey Designer. The Gallup Organization served as the local survey designer.

Name of the Panel (if applicable). Gallup Panel

Number of Data Collections. 1

Population Represented. Its target population consisted of the general non-institutionalized population aged 18 and older in the United States.

Interview Language. English

Sampling Method (selection of respondents, panel recruitment). Most panel members were recruited from Gallup Nightly tracking, a nightly study of 1,000 Americans on various topics. The sample for Gallup Nightly tracking was drawn using random digit dialing, which included a stratification of 50% landline users and 50% cell phone users with additional quotas by region. Landline respondents were chosen at random within each household (based on which member had the most recent birthday). For the cell phone frame, the person who answered the phone was selected for the sample. At the end of the interview, the respondents were asked whether he or she would like to participate in future surveys. If the respondent answered affirmatively, he or she was asked to join the Gallup Panel.

For our study, a stratified random sample was drawn from the Gallup Panel. The sample was stratified by age, education, and race. Only panel members that agreed to participate in surveys via the web received an invitation via email.

Sampling Frame (non-coverage, off-liners included). Persons without an Internet connection were not included in the sample of this study.

Sampling Frame Supplier (if applicable). The sample was drawn by the Gallup organization.

Incentives. Respondents did not receive incentives for participation.

Reminder/Number of Contact Attempts. Two reminders were sent out to respondents.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 1970 respondents

Dates of Data Collection. The field period lasted from April 2, 2014, to April 10, 2014.

Weights. The base weight includes the probability of selection into the panel, and the post-stratification weight is based on age, gender, education, ethnicity, and region.

Response Rates. A total of 5,000 persons were invited and a total of 1,970 interviews could be obtained. The completion rate was 39.4% (see AAPOR RR6), and the cumulative response rate was 1.6% (AAPOR CUMRR).

USA (Knowledge Panel/TESS)

Survey Sponsor (who paid for the survey). The study was sponsored by Time-sharing Experiments for the Social Sciences (TESS, NSF Grant 0818839, Jeremy Freese and James Druckman, Principal Investigators).

Data Collection Organization (who collected the data). The Knowledge Panel, in which the survey questions of this study were implemented, was conducted by GfK Custom Research (KnowledgePanel was acquired by Ipsos in October 2018).

Local Survey Designer. The core team served as the local survey designer.

Name of the Panel (if applicable). Knowledge Panel

Number of Data Collections (if applicable). 1

Population Represented. The general target population was the general US-American population aged 18 and older.

Interview Language. English

Sampling Method (selection of respondents, panel recruitment). Before 2009, random digit dialing methodology was used to recruit panel members. After 2009, the sample frame of the U.S. Postal Service's Delivery Sequence File replaced the recruitment by telephone. Randomly sampled addresses were invited to join the Knowledge Panel through a series of mailings, which were also available in Spanish to account for the Hispanic population. Non-responders were approached by telephone if a phone number could be matched to the sampled address. The invited households could join the panel using various means (postal reply, calling a hotline, and logging in on a recruitment website).

For this study, persons were drawn from the pool of panel members using a probability proportional to size (PPS) weighted sampling approach.

Sampling Frame (non-coverage, off-liners included). Households in the sample that did not

have the necessary equipment to take part in the web-based survey received a netbook computer and free Internet service.

Sampling Frame Supplier (if applicable). The sample was drawn by GfK.

Incentives. Respondents were rewarded with points, which can be cashed in for cash or merchandise.

Reminder/Number of Contact Attempts. Non-respondents received one reminder on the third day of the field period.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 1,029 respondents

Dates of Data Collection. The field period lasted from April 7, 2014, until April 15, 2015.

Weights. The base weight includes the probability of selection into the panel, and the post-stratification weight is based on age, gender, education, ethnicity, region, income, and home Internet access.

Response Rates. A total of 1,666 persons were invited to participate in this survey, and 1,029 persons did so. The completion rate was 61.8% (see AAPOR RR6), and the cumulative response rate was 5.6% (AAPOR CUMRR).

Canada

Survey Sponsor (who paid for the survey). The data collection was funded by the College of Arts and Sciences at the University of Saskatchewan.

Data Collection Organization (who collected the data). The first data collection was conducted by Social Sciences Research Laboratories (SSRL) at the University of Saskatchewan, and the second data collection was conducted by the *Probit* organization.

Local Survey Designer. For both data collections, the local team at the University of Saskatchewan served as the local survey designer.

Name of the Panel (if applicable). *Probit* panel

Number of Data Collections (if applicable). 2

Population Represented. The target population was the general Canadian population aged 18 and older.

Language. English and French

Sampling Method (selection of respondents, panel recruitment). In the first data collection, persons were recruited via telephone using random digit dialing (RDD) and then asked to fill out a web questionnaire. Respondents were sampled proportionately for each province in Canada.

In the second data collection, the telephone recruitment was complemented by a random probability-based online panel (*Probit* panel). Respondents were sampled proportionately for each province in Canada. The panel was recruited using random digit dialing based on a blended land-line cell-phone frame, and includes more than 90,000 panel members. In Stage 1, respondents were contacted via interactive voice (IVR) response. In Stage 2, those who expressed interest in becoming a panel member via IVR were contacted by an interviewer who guided them through the recruitment process.

Sampling Frame (non-coverage, off-liners included). Persons without an Internet connection

were not included in the sample of this study.

Sampling Frame Supplier (if applicable). The sample was drawn by the SSRL.

Incentives. Respondents did not receive incentives for participation.

Reminder/Number of Contact Attempts. Two email reminders were sent during the online data collection period. And the number of contact attempts by telephone was limited to ten attempts.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 1,317 respondents (first data collection: 319 respondents, second data collection: 998 respondents)

Dates of Data Collection. The field period for the first data collection was from April 7, 2014, until April 15, 2015, and for the second data collection between July 4, 2014, and July 22, 2014.

Weights. The SSRL did not provide a weight.

Response Rates. Based on a recruitment rate of 4.6% and a completion rate of 89.6%, the response rate for the first data collection was 4.1% (see AAPOR CUMRR). For the second data collection, the completion rate was 6.8% (see AAPOR RR6). For the second data collection, the information necessary to calculate the cumulative response rate was not provided.

Denmark

Survey Sponsor (who paid for the survey). The study was sponsored by Aalborg University.

Data Collection Organization (who collected the data). The data was collected by the consultancy Rambøll Management Consulting on behalf of Aalborg University.

Local Survey Designer. The local team at Aalborg University served as the local survey designer.

Name of the Panel (if applicable). The data was collected in the context of the ISSP 14 module.

Number of Data Collections (if applicable). 1

Population Represented. The target population was the general non-institutionalized population aged 18 to 79 in Denmark.

Interview Language. Danish

Sampling Method (selection of respondents, panel recruitment). A simple random sample was drawn from the Central Population Register (CPR) in Denmark.

Sampling Frame (non-coverage, off-liners included). Respondents in the sample that did not have the necessary equipment to take part in the web-based survey could take part via mail or telephone.

Sampling Frame Supplier (if applicable). The sample was provided by 'Forskerville' (A Public Danish Research Service Institution).

Incentives. As an incentive for their participation in the study, each respondent had the possibility to participate in a lottery; the prize was an Apple iPad.

Reminder/Number of Contact Attempts. In the first step, every person received an introduction letter via mail. The introduction letter included a link to a web survey and a unique personal code with which the survey could be activated. If respondents did not complete the web survey, they received a follow-up letter and were contacted via telephone and encouraged to participate in the

web questionnaire. Respondents also had the option to request a self-completion paper questionnaire and a telephone interview.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a “hard” probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability.

Survey Mode. A mixed methods design (online, mail, and telephone) was implemented to collect the data.

Sample Size. 1,325 respondents

Dates of Data Collection. The field period lasted from November 1, 2013, until January 3, 2014.

Weights. The data includes a post-stratification weight, which is based on age and gender.

Response Rates. A total of 2,499 persons were invited to participate in this survey, and 1,325 (1,273 online, 34 mail, and 18 telephone) persons did so. In addition, 63 partial completed interviews could be obtained. The response rate was 55.6% (see AAPOR RR6).

France

Survey Sponsor (who paid for the survey). The data collection was part of the Data, Infrastructure, and Methods of Investigation in the Social Sciences and Humanities (DIME-SHS) project, which was funded from the French National Agency for Research (ANR-10-EQPX-19-01). In addition, the recruitment of the ELIPSS pilot – in which the study was conducted – was partially funded by the French National Institute for Prevention and Health Education (INPES).

Data Collection Organization (who collected the data). The data collection was conducted in the framework of the ELIPSS (Étude Longitudinale par Internet Pour les Sciences Sociales) Panel.

Local Survey Designer. The local team at the Science Po served as the local survey designer.

Name of the Panel (if applicable). ELIPSS Panel

Number of Data Collections (if applicable). 1

Population Represented. The target population is the French population aged 18-75, residing in private households.

Interview Language. French

Sampling Method (selection of respondents, panel recruitment). The ELIPSS Panel was first implemented in 2012. The national statistical institute of France (INSEE) drew a stratified two-stage probability sample of 4,500 housing units from the rotating census. These samples were clustered in randomly drawn areas. At the third stage, one person was randomly selected from all adults in the household. Of the 4,500 housing units initially drawn from the target population, 1,039 persons were willing to take part in the monthly surveys.

Sampling Frame (non-coverage, off-liners included). Each panelist received a tablet equipped with wireless capability to participate in the self-administrated online questionnaires so that also respondents without Internet access were able to participate. Data is collected via the app pre-

installed on the tablet.

Sampling Frame Supplier (if applicable). The entire sample still in the panel at the time (i.e., 945 panelists) received the survey.

Incentives. The respondents did not receive any further incentive apart from the tablet provided to answer the questionnaires.

Reminder/Number of Contact Attempts. The ELIPSS Panel uses two types of reminders: on the one hand, automatic reminders by email, text message on the tablet and on the ELIPSS application; on the other hand, the panel uses follow-up calls to panelists who had been also non respondent to the previous waves. There were three automatic reminders and then up to three follow-up calls.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 835 respondents

Dates of Data Collection. The field work for this specific panel wave was done between April 3, 2014 and June 5, 2014.

Weights. The data includes a post-stratification weight, which is based on gender, age, nationality, and education.

Response Rates. For the panel wave that included the study questions, the completion rate was 88% (see AAPOR RR6) based on 945 invited persons and 835 completed interviews, and the cumulative response rate was 23.8% (AAPOR CUMRR).

Germany (German Internet Panel)

Survey Sponsor (who paid for the survey). The data collection was part of the German Internet Panel, which was funded in the framework of the Collaborative Research Center “Political Economy of Reforms” (SFB 884) by the German Research Foundation (DFG).

Data Collection Organization (who collected the data). The data was collected by the University of Mannheim.

Local Survey Designer. The local team at the University of Mannheim served as the local survey designer.

Name of the Panel (if applicable). German Internet Panel (GIP)

Number of Data Collections (if applicable). 2

Population Represented. The target population of the GIP consists of the general population aged 16-75 living in private households in 2012.

Interview Language. German

Sampling Method (selection of respondents, panel recruitment). The GIP is based on a three-stage probability sample. In the first stage, 250 PSUs, situated in 208 local administrative districts, were sampled. The sampling was stratified by state, governmental district, and level of urbanity. The households were drawn using a random route approach with a random starting point in each PSU with a separate listing of households. A total of 5,500 households were obtained in this sampling procedure. All household members were invited to join the panel. Computer-assisted face-to-face interviews were employed to recruit participants. A total of 1,603 panel members were willing to take part in the panel after the recruitment interview.

Sampling Frame (non-coverage, off-liners included). Households without access to the Internet and/or a computer received the necessary equipment and support.

Sampling Frame Supplier (if applicable). The entire sample received the survey.

Incentives. Respondents received 4€ for participation in each wave in addition to an annual bonus of 5€ or 10€ depending on the regularity of their participation.

Reminder/Number of Contact Attempts. Panel members received an email with the invitation to take part in the online survey. Panel members, who did not complete the survey, received more reminders per email and an additional reminder via telephone.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 1,063 respondents in Wave 8 and 1,064 respondents in Wave 9

Dates of Data Collection. The data were collected in the period from November 1, 2013 to November 30, 2013 for the first set of questions, and the data which included the second set of questions were collected in the period from January 1, 2014 to February 1, 2014.

Weights. The provided data set does not include a weight.

Response Rates. The completion rate for the first survey (Wave 8) was 66.9% based on 1,073 completed interviews (see AAPOR RR6), and the cumulative response rate was 12.4% (AAPOR CUMRR). The completion rate for the second survey (wave 9) was 66.4% based on 1,064 completed interviews (see AAPOR RR6), and the cumulative response rate was 12.3% (AAPOR CUMRR).

Germany (GESIS Panel)

Survey Sponsor (who paid for the survey). The data collection was part of the GESIS Panel, which was funded by the German Federal Ministry of Education and Research (BMBF).

Data Collection Organization (who collected the data). The data was collected by GESIS – Leibniz Institute for the Social Sciences.

Local Survey Designer. The local team at the GESIS served as the local survey designer.

Name of the Panel (if applicable). GESIS Panel

Number of Data Collections (if applicable). 1

Population Represented. The target population encompasses the German-speaking population aged between 18 and 70 years at the time of recruitment and permanently residing in Germany.

Interview Language. German

Sampling Method (selection of respondents, panel recruitment). The approx. 4,900 panelists were recruited offline in 2013, based on a random sample drawn from municipal population registers. A two-stage probability sampling scheme was employed: on the first stage sampling of municipalities, on the second stage sampling of individuals. The sampled individuals were contacted by an interviewer at their homes to conduct a personal recruitment interview.

Respondents willing to participate in the panel were asked in the recruitment interview whether they would like to participate online using a self-administered web survey or offline by filling out a paper questionnaire.

Sampling Frame (non-coverage, off-liners included). Respondents without Internet access participated via mail.

Probe. For all questions experiments no probe was used.

Sampling Frame Supplier (if applicable). The entire sample received the survey.

Incentives. Each invited panelist received an unconditional incentive of 5€ in the advance letter

and the offline respondent received the paper questionnaire in this letter as well.

Reminder/Number of Contact Attempts. For this study, independently from the survey mode in which the survey was conducted, all participants were invited per mail, and online respondents received an additional invitation email. Respondents invited per mail did not receive any reminders; online respondents received two reminders.

Probe. For all questions besides experiments on no opinion response options a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Mixed-mode survey (mail and web)

Sample Size. 4,298 respondents

Dates of Data Collection. The data were collected in the period from February 21, 2014 to April 14, 2014.

Weights. The provided data set does not include a weight.

Response Rates. A total of 4,298 panel members participated in the survey. The completion rate was 86.4% based on 4,221 completed interviews (see AAPOR RR6). The recruitment rate was 35.5% and the profile rate 64.5%. This resulted in a cumulative response rate of 19.8% (AAPOR CUMRR).

Iceland

Survey Sponsor (who paid for the survey). The data collection was funded by the Social Science Research Institute (SSRI) at University of Iceland.

Data Collection Organization (who collected the data). The data was collected by the SSRI at the University of Iceland.

Local Survey Designer. The local team at the University of Iceland served as the local survey designer.

Name of the Panel (if applicable). SSRI Internet panel

Number of Data Collections (if applicable). 1

Population Represented. The target population was the general population aged 18 and above.

Interview Language. Icelandic

Sampling Method (selection of respondents, panel recruitment). The SSRI Internet panel is based on a simple random sample drawn from the National Population Register which was provided by Registers Iceland. The recruitment was done via telephone interviews between 2010 and 2013 through different studies. Some studies had the sole purpose of recruiting panelists, while others were substantial CATI surveys. These interviews were concluded with an invitation to take part in the online panel. A stratified random sample of 4,987 individuals from the SSRI Internet Panel was invited to take part in this study. The sample was stratified by gender, age, and residence to reflect the composition of the Icelandic population in the best possible way (population information was acquired from Statistics Iceland).

Sampling Frame (non-coverage, off-liners included). Persons without an Internet connection were not included in the sample of this study.

Sampling Frame Supplier (if applicable). The entire sample received the survey.

Incentives. Respondents had the chance to win lottery prizes, which were gift certificates worth

approx. 70€

Reminder/Number of Contact Attempts. Respondents received an invitation to participate in the survey and reminders via email.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 4,987 respondents

Dates of Data Collection. The data collection was carried out between November 7, and November 27, 2013.

Weights. The data includes a post-stratification weight, which is based on gender, age, education, and resistance.

Response Rates. A total of 4,987 individuals were invited to take part in the survey, and 3,141 interviews could be obtained. The completion rate was 62.4% (see AAPOR RR6). The recruitment rate was 35.9%. This resulted in a cumulative response rate of 22.4% (AAPOR CUMRR).

Netherlands

Survey Sponsor (who paid for the survey). The data collection was part of the LISS Panel, which was funded by the Netherlands Organisation for Scientific Research (NOW).

Data Collection Organization (who collected the data). The data collection was conducted by the research institute CentERdata at the Tilburg University, the Netherlands.

Local Survey Designer. The core team served as the local survey designer.

Name of the Panel (if applicable). LISS (Longitudinal Internet Studies for the Social sciences) Panel

Number of Data Collections (if applicable). 1

Population Represented. The target population for the LISS Panel was the Dutch-speaking population permanently residing in the Netherlands aged 16 years or older.

Interview Language. Dutch

Sampling Method (selection of respondents, panel recruitment). Sampling units of the LISS panel were not individuals but households. The address frame for the sampling procedure was provided by Statistics Netherlands using a 10% sample from the population register GBA (Gemeentelijke Basisadministratie). For each address in the sample, a contact centre company which was part of the TNT post searched for the telephone number. This included landline numbers only. A single random sample of households was drawn without any stratification. The initial recruitment took place in 2007. Refreshment sample were drawn in 2009 and 2011.

Households were contacted first with an announcement letter. Next, respondents were contacted by an interviewer in a mixed mode design. Households for which a telephone number was available were contacted via telephone (CATI). The remaining households were visited by an interviewer and a face-to-face recruitment interview was conducted. After the short recruitment interview, all members of the household were invited to join the panel. Willing participants in

possession of the necessary equipment received a confirmation email, a letter with a login code, an information booklet, and a reply card and others were provided the technical equipment.

Sampling Frame (non-coverage, off-liners included). Participants without the necessary technical equipment to become panel members were loaned equipment to provide access to the Internet via a broadband connection.

Sampling Frame Supplier (if applicable). The sample was drawn by the research institute CentERdata at the Tilburg University.

Incentives. For each completed hour of interviews with the LISS Panel, respondents received 15€ that was paid quarter-yearly via bank transfer.

Reminder/Number of Contact Attempts. Respondents were invited by email to participate in the self-administered web survey. Respondents received two additional email reminders.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web Survey

Sample Size. 2,796 respondents

Dates of Data Collection. The field time of the survey that included the study questions was between January 6, 2014, and January 28, 2014.

Weights. The provided data set does not include a weight.

Response Rates. A total of 2,796 household members were selected for this study, of which 2,257 completed the interview. The completion rate was 80.6% (see AAPOR RR6). The information necessary to calculate the cumulative response rate was not provided by the LISS Panel.

Norway

Survey Sponsor (who paid for the survey). The Norwegian data was collected as part of the Citizen Panel. In addition to the UNI Rokkan Centre, the establishment of the Citizen Panel was financed by The Department of Comparative Politics, The Department of Administration and Organization Theory, The Department of Information Science and Media Studies and The Department of Sociology, all at the Social Science Faculty at the University of Bergen.

Numerous researchers and administrators from these organizations are involved in the operation and development of the citizen panel. Research projects financed by The Research Council of Norway also contribute to the financing of the panel, as do the University of Bergen and the Bergen Research Foundation.

Data Collection Organization (who collected the data). Ideas2Evidence was responsible for recruiting participants and sending out and receiving the surveys. The surveys and the panel are administered through the web-based research software Conformat.

Local Survey Designer. The local team at the University of Bergen served as the local survey designer.

Name of the Panel (if applicable). Norwegian Citizen Panel

Number of Data Collections (if applicable). 2

Population Represented. The target population encompassed the general population who had access to the Internet aged between 18 and 95.

Interview Language. Norwegian

Sampling Method (selection of respondents, panel recruitment). In the first wave in 2013, a national random sample of 25,000 individuals was drawn from the “National Population Register” (NPR) provided by the Norwegian Tax Agency. In the recruitment process, each person considered in the sample received a postal notification with information about the project and

information on how to participate. The invitation to join the panel, in general, was complemented with the invitation to participate in the first wave. A refreshment sample of additional 25,000 potential participants was drawn in the third wave, using the same sampling procedure and methodology as in the first wave in 2014.

Sampling Frame (non-coverage, off-liners included). Persons without an Internet connection were not included in the sample of this study. Individuals above the age of 95 were excluded from the sample. In wave 3 this amounted to 72 respondents, thereby leaving a net sample of 24,928 respondents before the recruitment process started.

Sampling Frame Supplier (if applicable). The entire sample received the survey.

Incentives. Participation was unpaid but a lottery on a travel gift card valued 25,000 NOK (about \$2900) has been included in each survey round as an incentive.

Reminder/Number of Contact Attempts. Existing panel members were invited by email to participate in the self-administered web survey, and new participants (Wave 3) were invited via mail to participate in the web survey. Afterward three reminders were sent (For the new recruitment in Wave 3: First, reminder postcard or reminder email to respondents that had registered their email address but not completed the survey; second, reminder text message to respondents with a registered phone number; third, reminder phone call to a subset of 2,000 that had yet to answer the survey. For existing panel members in Wave 3: First, E-mail or text message to respondents with a registered phone number; second, reminder E-mail. For existing Panel members in Wave 4: Three reminder emails, fourth reminder text messages to respondents with a registered phone number.)

Probe. For all questions no probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 8,515 respondents in Wave 3 and 6,297 respondents in Wave 4

Dates of Data Collection. Wave 3 (first set of questions) was a recruitment wave and was conducted in the period between October 13, 2014, and November 27, 2014. Wave 4 (second set of questions) was a regular panel wave and conducted using web survey interviews in the period between March 9 and April 8, 2015.

Weights. The data includes a post-stratification weight, which is based on gender, age, education, and region.

Response Rates. In Wave 3, a total of 24,928 persons were invited to take part in the survey, and 5,453 complete interviews could be obtained (NB: Number of total survey respondents was used when calculating response rates: 5,588). Besides the new recruitment from Wave 3, respondents previously recruited in Wave 1 were invited to respond to the Wave 3 questionnaire. From 4,833 individuals contacted, 2,927 responded. In sum, 29,761 individuals were contacted for the Wave 3 questionnaire (whether new recruitment or existing panel members); of these 8,515 responded, which resulted in a response rate of 28.6%.¹³

In Wave 4, although a total of 10,509 persons was invited to take part in the survey, in practice only 9,125 had taken part in at least one of the three preceding rounds; non-participants were thus considered as having withdrawn. Given a total of 6,297 interviews, the completion rate was thus 69% (see AAPOR RR6).

The NCP does not operate with outright “profile interviews.” Thus, cumulative response rates should be calculated directly as the relationship between all responses received and all

¹³ Note, that the response rate of 28.6% may be somewhat confusing, as it represents a compound of a study completion rate (the share of existing survey panel members completing the wave) and recruitment rate (the share of individuals invited by postal mail in connection with Wave 3 to take part in the panel actually completing the Wave 3 questionnaire). The study completion rate for the existing members recruited in wave 1 was 60.6%; the response rate for the new invitees in wave 3 was 23.1%.

individuals contacted. In Wave 3, this rate was 17% (AAPOR CUMRR); in Wave 4 it was 13% (AAPOR CUMRR).

Sweden

Survey Sponsor (who paid for the survey). The funding of LORE, which sponsored the study, was provided by the University of Gothenburg.

Data Collection Organization (who collected the data). The data of the Swedish subsample were collected by the Laboratory of Opinion Research (LORE) based at the University of Gothenburg.

Local Survey Designer. The local team at the University of Gothenburg served as the local survey designer.

Name of the Panel (if applicable). The Swedish Citizen Panel

Number of Data Collections (if applicable). 1

Population Represented. The target population was the Swedish population, which was defined as Swedish citizens and foreign nationals residing in Sweden over a year whose age was between 18 and 70 years.

Interview Language. Swedish

Sampling Method (selection of respondents, panel recruitment). The sample of the study consisted only of probability-based recruitment from population samples. The sample frame for the random population sample was the Swedish population register which was provided by the Swedish Tax Agency. For this study, a single recruitment cohort was used, 23,500 persons were invited to register for participation in the panel. This sample was selected using simple random sampling and each selected individual was contacted by a mailed invitation to become a panel participant. This recruitment resulted in 2,582 new panel members. LORE did not use separate steps for initial registration and a profile survey, rather those who joined the panel completed the profile survey at that same occasion as the initial consent to join the panel. The combined recruitment and profile rate for this recruitment cohort was 11.0 % (2,582/23,500). This can also

be formulated as $RECR \times PROR = 11\%$. For this specific study, 2,500 panel members from this recruitment cohort of 2,582 were randomly selected and invited via e-mail to answer the questions of this specific study.

Sampling Frame (non-coverage, off-liners included). The national population register served as a sampling frame for the panel recruitment. All residents of Sweden aged 18-70 were included in the sampling frame. Persons without access to an Internet connection could not become panel members and were thus not included in the sample of this study. At the time, approximately 92% of the population had internet access. For this specific study, the recruitment cohort described above, consisting of 2,582 new panel members, were included in the sampling frame.

Sampling Frame Supplier (if applicable). For the panel recruitment the Swedish Tax Agency served as sampling frame supplier. For the sampling to this specific study the local team at the Laboratory of Opinion Research at the University of Gothenburg managing the web panel supplied the sampling frame and drew the sample.

Incentives. Respondents did not receive any incentives for their participation in the survey.

Reminder/Number of Contact Attempts. The persons in the sample received an email with a unique direct link to the survey. One reminder was sent to those who did not finish the survey within twelve days using the initial invitation.

Probe. A soft probe asking respondents to please answer the question was used in all questions except no opinion response options and demographic and voting intention questions, which did not have a probe.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 2,500 respondents

Dates of Data Collection. The fieldwork was carried out between March 6 and April 7, 2014.

The reminder was sent out March 18.

Weights. The provided data set included a weight constructed on Swedish demographic statistics on sex, age and education.

Response Rate. Of the 2,500 persons invited to take part in the survey, 1,770 started the survey, and 1,725 provided complete responses (over 80 percent of eligible items are completed). There were 14 partial responses (answered between 50 and 80 percent of the questions given). The completion rate (COMR) was 69.6 % (see AAPOR RR6), and AAPOR RR5 69.0 %. The cumulative response rate (AAPOR CUMRR) was 7.7% ($11.0 \% \times 69.6\%$).

Taiwan

Survey Sponsor (who paid for the survey). The data collection was funded by the Center for Humanities and Social Sciences, Academia Sinica.

Data Collection Organization (who collected the data). The data of the Taiwanese subsample were collected by the Center for Survey Research (CSR) at the Center for Humanities and Social Sciences, Academia Sinica.

Local Survey Designer. The local team at the Academia Sinica served as the local survey designer.

Name of the Panel (if applicable). The sample was collected based on follow-up studies of multiple previous studies.

Number of Data Collections (if applicable). 2

Population Represented. The target population was the Taiwanese population between 18 and 70 years.

Interview Language. Chinese

Sampling Method (selection of respondents, panel recruitment). The sampling procedure was based upon email addresses that were provided at the end of different previously conducted random probability surveys: Nutrition and Health Survey (2009, face-to-face), Taiwan Genomic Survey (2009, face-to-face), Taiwan Social Change Survey (2011, face-to-face), Taiwan Weather Change Survey (2012 and 2013, telephone), and Taiwan Panel Study of Family Dynamics (2007-2014, face-to-face). All these surveys were conducted by the CSR. Additionally, email addresses from the Taiwan Panel Study of Family Dynamics were added in the second round of data collection.

Sampling Frame (non-coverage, off-liners included). Persons without an Internet connection were not included in the sample of this study.

Sampling Frame Supplier (if applicable). All participants that provided email addresses were invited.

Incentives. Participants in the survey were eligible for a raffle draw with prizes from approximately 5€(NT\$ 200) to approximately 140€(NT\$ 5,000).

Reminder/Number of Contact Attempts. Those participants of the surveys above who provided an email address were contacted and invited to take part in the self-administered web survey.

Four reminders were sent to those not participating.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Web survey

Sample Size. 326 respondents in the first data collection and 463 respondents in the second data collection.

Dates of Data Collection. The first round of data collection was carried out from July 21, 2014, to August 10, 2014, and the second round between February 24 and March 31, 2015.

Weights. The provided data set does not include a weight.

Response Rates. The first round of data collection led to 327 completed interviews out of an initial sample of 2,315 invited individuals. The completion rate was 15.3% (see AAPOR RR6). The cumulative response rate was 2.3% (AAPOR CUMRR). In the second round of data collection, 1,419 panelists were initially invited to take part in the survey, of which 463 interviews could be obtained. The completion rate was 37% (see AAPOR RR6). The cumulative response rate was 14.9% (AAPOR CUMRR). The overall cumulative response rate was 4.6% (AAPOR CUMRR).

The United Kingdom

Survey Sponsor (who paid for the survey). The British data collection was part of the Understanding Society Panel. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments.

Data Collection Organization (who collected the data). The data of the British sample were collected by the University of Essex, NatCen Social Research, and Kantar Public.

Local Survey Designer. The local team at the University of Essex served as the local survey designer.

Name of the Panel (if applicable). Understanding Society Innovation Panel

Number of Data Collections (if applicable). 1

Population Represented. The target population of the panel was the general population aged 16 and older living in private households in England, Scotland, and Wales.

Interview Language. English

Sampling Method (selection of respondents, panel recruitment). In the first wave of the Innovation Panel, a three-stage random sample was drawn. At the first stage, a systematic random sample of 120 postcode sectors from the Postcode Address File was drawn with probability proportional to population size after ordering by Government Office Region, the percentage of household heads classified as National Statistics Socio-Economic Classification categories 1 and 2 (non-manual), and population density. At the second stage, in each PSU 23 addresses were drawn as a systematic random sampling. At the final stage of sampling, interviewers conducted face-to-face interviews with all persons belonging to the target population at the time the interviewer conducted the recruitment interview. In Wave 4, an additional 960 addresses and in Wave 7, an additional 1,560 addresses from the original PSUs were added using systematic random sampling as refreshment samples.

Sampling Frame (non-coverage, off-liners included). Off-liners could take part in the face-to-face interviews.

Sampling Frame Supplier (if applicable). The entire panel was invited.

Incentives. An experiment concerning conditional and unconditional incentives was conducted. 50% of the households were issued to standard unconditional incentives. The incentives differed in their amount and their conditionality. Respondents recruited in wave 7 were allocated to three random groups, receiving either £10, £20 or £30 unconditionally.

Reminder/Number of Contact Attempts. Adults in households allocated to the web design group were sent an advance letter and an email if they provided an address. They received a URL and a unique log-in code. Several reminders were sent via email and eventually a reminder letter. Web respondents could join the face-to-face group if they did not complete the questionnaire online three weeks after the initial letter was sent.

Probe. For all questions except experiments on no opinion response options, which did not have a probe, a soft probe asking respondents to please answer the question was used.

Type of Sample (probability/non-probability). Probability

Survey Mode. Mixed-mode survey (online and face-to-face)

Sample Size. 2,262 respondents

Dates of Data Collection. The field period ran from May 21, 2014, to October 19, 2014.

Weights. The base and post-stratification weight is based on sex, age, and region.

Response Rates. A total of 5,415 adults were invited, and a total of 2,262 individual interviews with adults could be obtained. The individual response rate was 42.5% (see AAPOR RR6), and the response rate of the refreshment sample was 29.2% (AAPOR RR1). The cumulative response rate on the household level was 36.2% (AAPOR CUMRR).

Appendix B. Translated Question Wordings

Question Wording Experiment

Source questions in English (Canada, U.S., and UK)

There are some people who are against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed the freedom to speak, or not?

Yes, allowed to speak

No, not allowed

There are some people who are against all churches and religion. If such a person wanted to make a speech in your (city/town/community) against churches and religion, should he be allowed to speak, or not?

Yes, allowed to speak

No, not allowed

Chinese (Taiwan)

有些人反對所有的宗教。如果有這樣的人，想在您住的地區發表反對宗教的言論，是否應該允許他／她有發言的自由？

是，應該允許

否，不應該允許

有些人反對所有的宗教。如果有這樣的人，想在您住的地區發表反對宗教的言論，是否應該允許他／她發言？

是，應該允許

否，不應該允許

Danish (Denmark)

Nogle personer er imod alle former for kirker og religioner. Hvis én af disse personer ville tale offentligt i dit lokalområde imod kirker og religioner, skulle vedkommende så have ret til at bruge sin ytringsfrihed og tale eller ej?

Ja, vedkommende skulle have ret til at tale

Nej, vedkommende skulle ikke have ret til at tale

Nogle personer er imod alle former for kirker og religioner. Hvis én af disse personer ville tale offentligt i dit lokalområde imod kirker og religioner, skulle vedkommende så have ret til at tale eller ej?

Ja, vedkommende skulle have ret til at tale

Nej, vedkommende skulle ikke have ret til at tale

Dutch (Netherlands)

Er zijn sommige mensen die tegen alle kerken en religies zijn. Als zo iemand een toespraak zou willen houden in uw stad/dorp/gemeente tegen kerken en religies, zou hij of zij dan de vrijheid moeten krijgen om te spreken of niet?

Ja, moet mogen spreken

Nee, moet niet mogen spreken

Er zijn sommige mensen die tegen alle kerken en religies zijn. Als zo iemand een toespraak zou willen houden in uw stad/dorp/gemeente tegen kerken en religies, zou hij of zij dan moeten mogen spreken of niet?

Ja, moet mogen spreken

Nee, moet niet mogen spreken

French (France)

Il y a certaines personnes qui sont contre tous les cultes et la religion. Si une telle personne voulait faire un discours dans votre commune contre les cultes et la religion, devrait-elle avoir la liberté de s'exprimer ou non?

Oui, elle devrait pouvoir s'exprimer

Non, elle ne devrait pas pouvoir

Il y a certaines personnes qui sont contre tous les cultes et la religion. Si une telle personne voulait faire un discours dans votre commune contre les cultes et la religion, devrait-elle pouvoir s'exprimer ou non?

Oui, elle devrait pouvoir s'exprimer

Non, elle ne devrait pas pouvoir

French (Canada)

Il y a des gens qui sont contre toutes les églises et la religion. Si une telle personne voulait faire un discours dans votre (ville / commune / communauté) contre les églises et la religion, est-ce qu'il devrait avoir la liberté de parler, ou non?

Oui, le droit de parler

Non, pas permis

Il y a des gens qui sont contre toutes les églises et la religion. Si une telle personne voulait faire un discours dans votre (ville / commune / communauté) contre les églises et la religion, devrait-il avoir l'autorisation de parler, ou non?

Oui, le droit de parler

Non, pas permis

German (Germany)

Es gibt Menschen, die gegen alle Kirchen und Religion sind. Wenn so jemand in Ihrer Stadt/Gemeinde eine Rede gegen Kirchen und Religion halten wollte, sollte ihr/ihm die Freiheit zugestanden werden zu sprechen oder nicht?

Ja, zugestanden werden zu sprechen
 Nein, nicht zugestanden werden

Es gibt Menschen, die gegen alle Kirchen und Religion sind. Wenn so jemand in Ihrer Stadt/Gemeinde eine Rede gegen Kirchen und Religion halten wollte, sollte ihr/ihm zugestanden werden zu sprechen oder nicht?

Ja, zugestanden werden zu sprechen
 Nein, nicht zugestanden werden

Icelandic (Iceland)

Sumir eru á móti kirkjunni og öllum trúarbrögðum. Ef einstaklingur með slíkar skoðanir myndi vilja halda ræðu í (þinni borg/ þínum bæ/þínu samfélagi) gegn kirkjunni og öllum trúarbrögðum, ætti að veita einstaklingnum frelsi til að tala, eða ekki?

Já, ætti að fá að tala
 Nei, ætti ekki að fá að tala

Sumir eru á móti kirkjunni og öllum trúarbrögðum. Ef einstaklingur með slíkar skoðanir myndi vilja halda ræðu í (þinni borg/ þínum bæ/þínu samfélagi) gegn kirkjunni og öllum trúarbrögðum, ætti einstaklingurinn að fá að tala eða ekki?

Já, ætti að fá að tala
 Nei, ætti ekki að fá að tala

Norwegian (Norway)

Noen mennesker er imot alle trossamfunn og all religion. Hvis en slik person ønsket å tale imot trossamfunn og religion i (din by/ditt hjemsted/ditt nærmiljø), burde han/hun fått lov til å bruke ytringsfriheten eller ikke? Ja, burde fått lov til å tale

Nei, burde ikke fått lov

Noen mennesker er imot alle trossamfunn og all religion. Hvis en slik person ønsket å tale imot trossamfunn og religion i (din by/ditt hjemsted/ditt nærmiljø), burde han/hun fått lov til å tale eller ikke?

Ja, burde fått lov til å tale

Nei, burde ikke fått lov

Swedish (Sweden)

Det finns vissa personer som är emot alla kyrkor och religioner. Om en sådan person skulle vilja hålla ett tal mot kyrkor och religion på din ort, bör personen tillåtas friheten att tala eller inte?

Ja, tillåtelse att tala

Nej, inte tillåtelse

Det finns vissa personer som är emot alla kyrkor och religioner. Om en sådan person skulle vilja hålla ett tal mot kyrkor och religion på din ort, bör personen tillåtas att tala eller inte?

Ja, tillåtelse att tala

Nej, inte tillåtelse

Appendix C. Measurement of the education in each country

Canada

Low education: Some elementary school, completed elementary school, some secondary/high school, completed secondary/high school, some technical or community college

Medium education: Completed technical or community college, some university

High education: Bachelor's degree, master's degree, professional degree (e.g., law degree, medical degree), doctorate

Denmark

Low education: 7 years primary school or shorter, 8 years primary school, 9 years primary school, secondary, 10 years or similar, gymnasium general, gymnasium technical, other school education, basic vocational and apprenticeship, other vocational education, other completed vocational education

Medium education: Short advanced education, less than 3 years, middle range advanced, 3-4 years

High education: Further advanced, more than 4 years

Germany (GESIS and GIP)

Low education: 9 years primary school and shorter

Medium education: 10 years secondary school, 11 years secondary school

High education: professional degree, bachelor's degree, master's degree, doctorate

France

Low education: Level VI – without diploma or certificate of general education, level V – certificate of professional competence or technical school certificate

Medium education: Level IV – general, technological or professional baccalaureate, level III – two-year post-baccalaureate diploma (technology university degree, advanced technical certificate)

High education: Level II and I: graduate or post-graduate (bachelor's degree, master's degree, and doctorate)

Iceland

Low education: Compulsory education

Medium education: Secondary education

High education: University education

Japan

Low education: Middle

Medium education: High

High education: University/college

Netherlands

Low education: Primary school, junior high school

Medium education: Senior high school, junior college

High education: College, university

Norway

Low education: No completed education, primary school, 7 year elementary school, county college certificate, 6th form college/end of formal schooling

Medium education: Preliminary course university/college without credits, certificate for additional further education, university/college (less than 3 years but at least 2 years)

High education: Completed 3-4 years of college/university education, completed 5-6 years of college/university education, PhD

Portugal

Low education: None, incomplete primary education, 4 years schooling, 6 years schooling, 9 years schooling

Medium education: 12 years schooling (high school, secondary school)

High education: Complete polytechnic/post-secondary, complete trade/vocational school, university bachelor and higher)

Sweden

Low education: Not completed elementary school, elementary school, less than 3 years high school, 3 or more years high school

Medium education: less than 3 years studies after high school (not college/university), 3 years more year's studies after high school (not college/university), less than 3 years university/college

High education: 3 years or more university/college, PhD

Taiwan

Low education: Junior high school, vocational junior high school, senior high school, vocational senior high school, two-year junior college, three junior college, five-year junior college, military/police two-year junior college, open junior college

Medium education: Open University, military/police College, institute of technology

High education: University, master's degree, Doctoral degree

United Kingdom

Low education: No qualification, other qualification, GCSE etc.

Medium education: A-level etc., other secondary degree

High education: University degree

United States of America (Gallup and TESS)

Low education: Less than high school complete, 8th grade or less, 9th-12th grade (no high school diploma), high school graduate

Medium education: Trade/technical/vocational training, some college

High education: College graduate, postgraduate work

Appendix D. Question Wording – ESS Questions

English (UK)

Freedom

It is important to him to make his own decisions about what he does. He likes to be free and not depend on others.

Very much like me, like me, somewhat like me, a little like me, not like me, not like me at all

Religiosity

Regardless of whether you belong to a particular religion, how religious would you say you are?

(0) not at all religious, (10) very religious

Danish (Denmark)

Freedom

Det er vigtigt for ham selv at bestemme, hvad han foretager sig. Han kan godt lide at være selvstændig og uafhængig af andre

Ligner mig meget, ligner mig, ligner mig i nogen grad, ligner mig lidt, ligner mig ikke, ligner mig slet ikke

Religiosity

Uanset om du tilhører en bestemt religion eller ej, hvor religiøs vil du sige, at du er?

(0) Slet ikke religiøs, (10) Meget religiøs

Dutch (Netherlands)

Freedom

Het is belangrijk voor hem om zelf beslissingen te nemen over wat hij doet. Hij wil graag vrij en onafhankelijk van anderen zijn.

Lijkt heel erg veel op mij, Lijkt op mij, Lijkt enigszins op mij, Lijkt weinig op mij, Lijkt niet op mij, Lijkt helemaal niet op mij

Religiosity

Ongeacht of u zichzelf nu wel of niet als lid van een bepaald geloof of kerkgenootschap beschouwt, hoe gelovig vindt u dat u bent?

(0) Helemaal niet gelovig, (10) Heel erg gelovig

French (France)

Freedom

C'est important pour lui de décider tout seul de ce qu'il va faire. Il aime être libre et ne pas dépendre des autres.

tout à fait comme moi, comme moi, plutôt comme moi, un peu comme moi, pas comme moi, pas du tout comme moi

Religiosity

Indépendamment de votre appartenance ou non à une religion, diriez-vous que vous êtes quelqu'un de croyant ?

(0) pas du tout croyant, (10) très croyant

*German (Germany)*Freedom

Es ist ihm wichtig, selbst zu entscheiden, was er tut. Er ist gerne frei und unabhängig von anderen.

Ist mir sehr ähnlich, ist mir ähnlich, ist mir etwas ähnlich, ist mir nicht ähnlich, ist mir überhaupt nicht ähnlich

Religiosity

Unabhängig davon, ob Sie sich einer bestimmten Religion zugehörig fühlen, für wie religiös würden Sie sich selber halten?

(0) überhaupt nicht religiös, (10) sehr religiös

*Icelandic (Iceland)*Freedom

Honum finnst mikilvægt að taka sínar eigin ákvarðanir. Hann vill vera frjálfs og óháður öðrum. Mjög líkt mér, Líkt mér, Nokkuð líkt mér, Svólítið líkt mér, Ekki líkt mér, Alls ekki líkt mér

Religiosity

Án tillits til þess hvort þú aðhyllist einhver sérstök trúarbrögð, hversu trúuð/trúaður myndirðu segja að þú sért?

(0) Alls ekki trúuð/trúaður, (10) Mjög trúuð/trúaður

*Norwegian (Norway)*Freedom

Det er viktig for ham å selv bestemme hva han skal gjøre. han liker å være fri og ikke avhengig av andre.

Veldig lik meg, Lik meg, Noe lik meg, Litt lik meg, Ikke lik meg, Ikke lik meg i det hele tatt

Religiosity

Uavhengig av om du tilhører en bestemt religion, hvor religiøs vil du si at du er?

(0) Ikke religiøs i det hele tatt, (10) Veldig religiøs

*Swedish (Sweden)*Freedom

Det är viktigt för honom att fatta sina egna beslut om vad han ska göra. Han vill vara fri och inte vara beroende av andra.

Är väldigt mycket som jag, Är som jag, Är till viss del som jag, Är lite grann som jag, Är inte som jag, Är inte alls som jag

Religiosity

Oavsett om du tillhör någon särskild religion eller inte, hur religiös skulle du säga att du är?

(0) Inte alls religiös, (10) Mycket religiös