Qual Quant (2014) 48:127–148
DOI 10.1007/s11135-012-9754-8

Item comparability in cross-national surveys: results from asking probing questions in cross-national web surveys about attitudes towards civil disobedience

Dorothée Behr · Michael Braun · Lars Kaczmirek · Wolfgang Bandilla

Published online: 10 August 2012

Abstract This article focuses on assessing item comparability in cross-national surveys by asking probing questions in Web surveys. The "civil disobedience" item from the "rights in a democracy" scale of the International Social Survey Program (ISSP) serves as a substantive case study. Identical Web surveys were fielded in Canada (English-speaking), Denmark, Germany, Hungary, Spain, and the U.S. A category-selection and a comprehension probe, respectively, were incorporated into the Web surveys after the closed-ended "civil disobedience" item. Responses to the category selection-probe reveal that notably in Germany, Hungary, and Spain the detachment of politicians from the people and their lack of responsiveness is deplored. Responses to the comprehension probe show that mainly in the U.S. and Canada violence and/or destruction are associated with civil disobedience. These results suggest reasons for the peculiar statistical results found for the "civil disobedience" item in the ISSP study. On the whole, Web probing proves to be a valuable tool for identifying interpretation differences and potential bias in cross-national survey research.

Keywords Probing · Web surveys · Mixed method · Comparability · Cross-national survey research

1 Introduction

1.1 Comparability needs in cross-national survey research

Large-scale cross-national survey projects, such as the World Value Surveys (WVS), the European Values Study (EVS), the International Social Survey Program (ISSP), the European Social Survey (ESS), and the Eurobarometer have established long time series, partly beginning in the 1970s and 1980s. A major goal of these surveys is the analysis of social change in a comparative perspective, which requires the continued invariant measurement of

D. Behr (✉) · M. Braun · L. Kaczmirek · W. Bandilla
GESIS – Leibniz Institute for the Social Sciences, P.O. Box 122155, 68072 Mannheim, Germany
e-mail: dorothee.behr@gesis.org

constructs across countries. However, countries can differ in the way the majority of the population interprets a question, even if the question seems well-translated. In addition, social change over time may lead to changes in item interpretation in a country. These aspects are a threat to the long-term validity of cross-national surveys because differences across countries and over time could be methodological artifacts rather than "real" differences between countries and time periods.

Good questionnaire design with diverse international input in form of expert review, pretesting, and the like (Harkness et al. 2010) can help prevent some of the problems mentioned above. However, even in the major cross-national surveys, pretesting across many different countries prior to questionnaire finalization is rare and so interpretation differences can never be fully ruled out. Moreover, appropriate questionnaire design cannot prevent items from changing their meaning across time or from taking on an unintended or differential meaning in a country which newly joins a survey program and which simply has to accept the (core) questionnaire as it is (Miller et al. 2011).

In view of these "threats", assessing data comparability is a necessary prerequisite of any substantive analysis of cross-national survey data. Although the application of data-analytic procedures is an appropriate means for detecting equivalence problems, they are not conducive to explaining the causes of incomparability. Knowledge of these causes, however, could serve to improve measurement instruments for future surveying. Even if measurement instruments cannot be changed, as is often the case in an ongoing time series, knowledge of the causes of incomparability can be usefully applied and inform substantive interpretation of the data.

1.2 Cognitive interviewing

A possible solution for detecting measurement artifacts and learning about different item interpretation across countries is to conduct cognitive interviewing. Cognitive interviewing helps to uncover the cognitive processes respondents use when answering a survey question. Respondents have several tasks to complete when answering a survey item: interpreting a question, generating an opinion, matching the opinion to a response category, and editing the response taking social desirability into consideration (e.g., Schwarz 1996; Strack and Martin 1987; Tourangeau et al. 2000). Surveys are likely to contain at least some questions which do not reflect or match social reality and issues of public debate. They may also include terminology that is not consistently understood by all respondents or that is too vague, thus opening up paths for a variety of possible interpretations. Here, different cultural contexts could lead to different item interpretations across countries. This may even be the case with items that seem well-translated.

There are two major cognitive interviewing techniques used in survey research. These are think-aloud where respondents verbalize their thoughts as they answer survey questions and verbal probing where interviewers ask follow-up questions to obtain additional information on responses (Beatty and Willis 2007; Willis 2005). Among the probing techniques, comprehension probing (such as "What does the term xy mean to you?") is a particularly suitable means to reveal country-specific interpretations of individual terms or phrases. Also category-selection probing (Prüfer and Rexroth 2005) can be regarded as an appropriate means to assess item comparability. With category-selection probing respondents are asked for their reasons for having selected a particular scale value for a closed question. The answers to category-selection probing allow to analyze the differentiations respondents make in the interpretation of an item. In particular "silent misinterpretations" (DeMaio and Rothgeb 1996)

can be detected where respondents seemingly do not have problems with the interpretation of an item but actually misinterpret its meaning.

1.3 Cognitive interviewing across countries

The present use of cognitive interviewing (techniques) in the cross-national context is restricted. First, these techniques typically serve to improve a comparative source questionnaire or a translation prior to fielding (e.g., Childs and Goerman 2010; Fitzgerald et al. 2011; Forsyth et al. 2007). Implementing cognitive techniques in a cross-national survey itself (similar to Schuman's "random probes", 1966) or conducting cognitive interviews "after the fact" to explain problems in the dataset is still rare. Recent exceptions, however, are the post-survey cognitive interviewing studies by Thrasher et al. (2011); Reeve et al. (2011) or Morren et al. (Forthcoming). Second, the comparative cognitive studies that are conducted mostly apply to different ethnic groups in only one country, mostly the U.S. (e.g., Goerman and Caspar 2010; Willis et al. 2008). While there are certainly challenges involved in organizing these studies in one country, it is even more difficult to organize them across countries, especially if comparability of results is the goal (Lee 2012). Experienced cognitive interviewers may not be available in all countries, and even if they were, it would be necessary to standardize procedures across countries for reasons of comparability. For instance, different house styles in recruiting respondents or different guidelines specifying the conduct of interviews would need to be harmonized, at least to some extent (Miller et al. 2011; Thrasher et al. 2011). Third, the number of cases per country in cross-national studies is usually too small to draw more generalizable conclusions on the differences between country-specific answer patterns (Thrasher et al. 2011). One of the few cross-national cognitive studies, in which seven countries participated and in which, among others, ESS questions were tested, conducted not more than ten interviews in three and not more than 20 in another two countries (Fitzgerald et al. 2011). Other large cross-national survey projects, such as the ISSP, do not implement comparative cognitive pretests at all. Instead, they leave it to individual countries whether such pretests are conducted and do not make an attempt to systematically compare the results across countries.

There is no compelling reason why methods for analyzing cognitive processes involved in the interpretation of items need to be restricted in the way they currently are. First, while assessment of survey questions ideally takes place before items are used in the field, much could be learned as well from post-survey evaluation and from targeting problems that show up in the final data set. Second, the application of cognitive interviewing techniques does not have to be based on personal interviews. The critical issues of (non-)availability of cognitive interviewers across countries or of harmonization of probing procedures (Thrasher et al. 2011) could be circumvented by implementing probing techniques in self-administered surveys. Third, the dominant assumption that low case numbers will be sufficient when using cognitive interviewing techniques is counterproductive. They do not permit the analysis of diverging argumentation patterns across countries or meaningful quantification of results. Also in this case, self-administered surveys seem a suitable option.

Combining all aspects mentioned above, we propose to implement probing questions in cross-national Web surveys to assess comparability of items. The general feasibility of asking probing questions on the Web has already been tested and confirmed by Behr et al. (2012) and Behr et al. (Forthcoming). In these studies, data collection and substantive analyses have been restricted to Germany only. This paper extends this research to the international arena and addresses the question to what extent probing techniques implemented in cross-national Web surveys allow to unravel different interpretation patterns across countries. Directly related

to this, we investigate the usefulness of different probing techniques (comprehension vs. category-selection) and investigate the contribution that each one can make to identifying comparability or non-comparability of items.

1.4 Case study: the "civil disobedience" item in the ISSP

To demonstrate the utility of implementing probing questions in cross-national Web surveys, we investigate why respondents answer the "civil disobedience" item from the 2004 ISSP Citizenship Module the way they do. The ISSP is an annual cross-national survey program, fielding general population surveys on topics that are relevant to the social sciences. Set up in 1984, the ISSP has grown from four to 48 member countries in 2011. The questionnaire design process is characterized by rigorous cross-cultural collaboration, which cannot preclude, however, that some items might eventually include bias of some sort or other. In the ISSP, the "civil disobedience" item is part of the following scale on rights of people in a democracy (ISSP Research Group 2004a):

There are different opinions about people's rights in a democracy. On a scale of one to seven, where one is not at all important and seven is very important, how important is it…

a. that all citizens have an adequate standard of living
b. that government authorities respect and protect the rights of minorities
c. that government authorities treat everybody equally regardless of their position in society
d. that politicians take into account the views of citizens before making decisions
e. that people be given more opportunities to participate in public decision-making
f. that citizens may engage in acts of civil disobedience when they oppose government action.

Results from the ISSP (ISSP Research Group 2004b) are displayed in Table 1. They show that Canada, Denmark, Germany (eastern/western), Hungary, Spain, and the U.S. have a similar high value for an index based on the items a–e (the index, indicating the mean score on these five items, ranges from 6.2 to 6.6). However, the means of the "civil disobedience" item shows a pattern that clearly distinguishes between two groups of countries. While the means for civil disobedience is lower than the index in all countries, it is particularly low for Canada, Denmark, and the U.S. (between 3.8 and 4.1). In Germany (eastern and western), Hungary, and Spain it ranges from 5.0 to 5.5.

This peculiar pattern could be explained in at least two ways. First, the differences may be due to country differences in trust in government. Donovan et al. (2008), for instance, find that the need to engage in civil disobedience is negatively correlated with trust in government. Trust in government depends on a variety of factors, among which the evaluation of democratic performance or that of governmental responsiveness. Equally, corruption plays its part in shaping trust in government.

Second, the peculiar pattern across countries might equally be due to country-specific interpretations of the term "civil disobedience". In Canada, Denmark, and the U.S., respondents might associate negative events or situations with the term and thus be less inclined to support civil disobedience. If the key term "civil disobedience" is indeed understood differently across countries, we would have to deal with a lack of cross-cultural validity. The high non-response rate for civil disobedience in the ISSP data set already suggests that something is problematic with the item (between 3.3 and 17.5 % across countries). The amount of non-response for the "civil disobedience" item has also led researchers to discard the item from analyses (e.g., Bolzendahl and Coffe 2008).

Table 1 ISSP and Web survey data for the "rights in a democracy" scale (standard deviations in parentheses)

Country	ISSP			Web surveys				
	Index (1–7)	Civil disobedience (1–7)	Civil disobedience non-response (%)	Index (1–7)	Civil disobedience (1–7)	Civil disobedience non-response (%)	Left–right (1–10/11)	Political interest (1–5)
Canada	6.2 (.81)	3.8 (2.0)	11.8	6.1 (.99)	4.3 (1.8)	7.0	5.7 (2.0)	2.7 (1.2)
Denmark	6.4 (.67)	4.1 (2.1)	16.2	6.0 (.84)	4.5 (1.7)	11.9	5.6 (2.3)	2.9 (1.1)
Germany (eastern)	6.4 (.61)	5.5 (1.5)	11.9	6.1 (.86)	5.5 (1.5)	6.9	5.2 (2.2)	3.1 (1.1)
Germany (western)	6.3 (.81)	5.2 (1.6)	17.5	6.1 (.81)	5.5 (1.5)	9.3	5.3 (2.0)	3.2 (1.1)
Hungary	6.2 (.90)	5.5 (1.7)	8.9	6.0 (.77)	5.9 (1.3)	3.7	6.2 (2.2)	2.8 (1.1)
Spain	6.6 (.63)	5.0 (2.0)	15.3	6.3 (.82)	5.2 (1.6)	7.8	5.4 (2.5)	3.0 (1.2)
U.S.	6.4 (.78)	4.0 (2.2)	3.3	6.0 (1.0)	4.5 (1.9)	5.0	6.0 (2.2)	3.0 (1.3)

Note: ISSP n = 8,281; Web surveys n = 3,695. The index includes the items a–e of the "rights in a democracy scale". The civil disobedience scale runs from one (not at all important) to seven (very important), and so does the index. The left–right self-placement (only Web survey) is based on eight experimental scales, combining ten and 11 points, DK and omission of DK as well as end-numbers only or a fully numbered scale. For the calculation of the mean, all of these scales were merged. For the political interest item (reverse coded, only Web survey), high values indicate high interest, low values indicate low interest

Table 2 Descriptive statistics of the dataset of the Web surveys

Country/region	n	Break-off (%)	Mean duration (min)
Canada	516	70 (12.0)	16:05
Denmark	537	81 (13.1)	15:40
Germany (eastern)	537	54 (9.1)	17:48
Germany (western)	507	40 (7.3)	15:51
Hungary	536	85 (13.7)	20:32
Spain	538	73 (12.0)	15:43
U.S.	524	61 (10.4)	15:57
Total	3,695	464 (11.4)	16:49

Our study investigates whether differences in the level of trust in government and/or different cognitive representations of the term civil disobedience contribute to country differences.

2 Data and methods

2.1 Data source

The data in this paper come from Web surveys that were conducted in Canada (English-speaking), Denmark, Germany (eastern and western Germany as separate regions), Hungary, Spain, and the U.S. in January 2011. Participants to these surveys were drawn from nonprobability access panels.[1] Our Web surveys targeted nationals aged 18–65. Quotas based on gender, age (18–30, 31–50, and 51–65), and education (lower education vs. higher education) were used to obtain a balanced, albeit not representative, sample. Thus, gender and education were similarly distributed, and the mean age ranged from 40.6 to 42.3. In total, 3,695 respondents across all countries completed the survey. The mean answer time (for all countries) was almost 17 min, and the mean break-off rate was 11.4 %. Table 2 provides descriptive statistics for the individual countries. The study presented here is part of a larger methods project that aims at detecting comparability problems in cross-national survey research through the development and assessment of Web probing.

2.2 Questionnaire

The questionnaire covered the topics of family (e.g., gender roles) and politics (e.g., attitudes towards rights of people in a democracy and towards immigrants) and mainly used items from the International Social Survey Program. The topical blocks were rotated to avoid sequence effects. In total, the questionnaire contained 36 closed-ended items. In addition, each respondent received eight open-ended probing questions, among which category-selection and comprehension probing. Soft edit checks were included with all items and probes: If respondents did not answer a question, a message was shown that reminded them of the importance of their answer. Respondents were then able to choose between giving an answer or skipping to the next question. The ISSP "rights of people in a democracy" scale, as presented above, was implemented with each item appearing on a separate screen. The response scale was end-point labeled and ranged from one "not at all important" to seven "very

[1] Among the countries of the study, probability-based Web panels to date only exist in the U.S. Therefore, we worked with providers of nonprobability access panels.

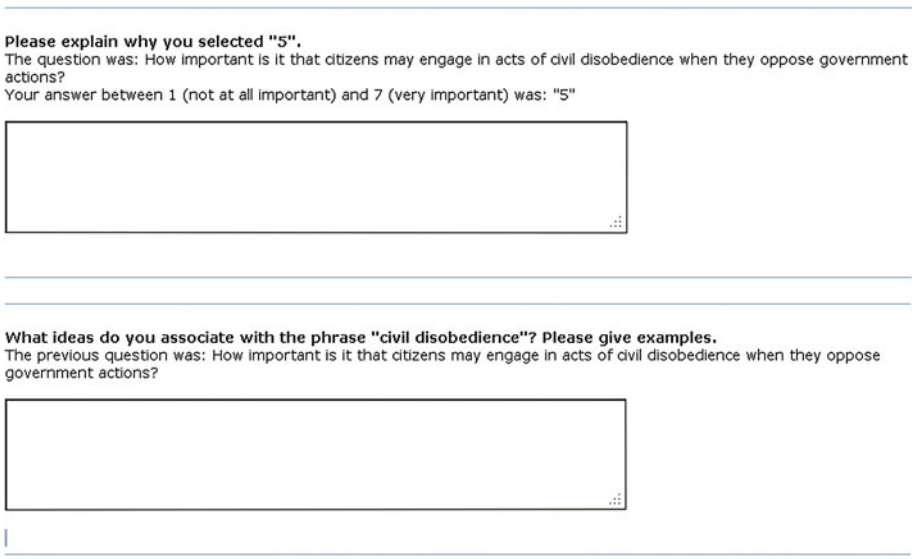

Fig. 1 Screenshots of the category-selection (*top*) and the comprehension probe (*bottom*)

important" (horizontal display). A "can't choose" category was explicitly offered to respondents. After the final item on civil disobedience, respondents were randomly assigned to either the screen with the comprehension probe ("What ideas do you associate with the phrase 'civil disobedience'? Please give examples.") or to the screen with the category-selection probe ("Please explain why you selected [1–7 or 'can't choose' inserted].") (see Fig. 1 for screenshots). Respondents thus received only one of these probes. Furthermore, within the politics section we inquired after respondents' interest in politics ("How interested are you in politics? Fairly interested, interested, some interest, little interest, not at all interested (1–5)"; reverse coded such that high values indicate high interest). In addition, we asked respondents to place themselves on the left-to-right scale ("Many people use the terms 'left' and 'right' to designate different political positions. We have a scale here that runs from left to right. Where would you position your own political views on this scale?"). The scale itself was randomly modified in eight experiments, including ten or 11 scale points, the DK option or not, and end-numbers only or the full set of numbers. Political interest and left-right orientation will serve as control variables in the analyses to assess the potential of bias we might have. After all, these variables are likely to be correlated with the civil disobedience item and might introduce bias if totally unrealistically represented in our Web surveys.

2.3 Translation of probe answers, development of the coding scheme, and coding

The Danish, Hungarian, and Spanish answers to the civil disobedience probes were translated into German by professional translators who had been briefed on the particularities of these texts as well as on translation and coding needs. The translators equally served as a point of contact for queries that came up during the coding process, e.g., on the scope of meaning of certain terms. Since we were able to code English and German answers in our research team, we did not commission any translation for these languages.

Two distinct substantive coding schemes were developed for the two types of probes owing to the different nature of responses that we obtained from respondents. Obviously,

explicitly asking for examples in the case of the comprehension probe led to a high number of respondents offering example lists such as "Picketing, rallying, speaking out publicly". Answers such as these were not common for the category-selection probe.

Multiple substantive coding was possible for each probe answer and further defined in the coding guidelines. In addition to substantive coding, non-response codes were assigned to answers. We considered the following to be non-response answers: (1) no text entry at all; (2) ?, —, letter combinations such as "fdg"; (3) don't knows; (4) refusals such as "n/a"; (5) other meaningless entries such as "it is like that", "it is not that important", "why not"; and (6) incomprehensible and therefore non-codeable answers such as "to an extent" or "against us" (see Holland and Christian 2009 for a similar non-response definition). We tried to build the substantive coding scheme for category-selection and comprehension probing in a way that similarities between both probing types became visible.

Inter-rater agreement based on 10 % of answers proved satisfactory for both the comprehension probe (between 79 and 88 % across countries) and the category-selection probe (between 75 and 92 % across countries). The low of 75 % was partly due to mismatches within the different non-response codes. Corrections to the dataset were made following the inter-rater assessments.

2.4 Analytical procedures

First, the Web survey data is compared to the ISSP data. We regard the replication of the country patterns for the index of the "rights in a democracy scale" (all countries having about the same mean for items a–e) and the "civil disobedience" item (two groups of countries with markedly different means) as a precondition for using the Web survey data to shed light on the ISSP data. Second, we present the substantive coding schemes for the two probe types. Third, we present the distributions of the codes across countries. The significance of country differences is assessed using logistic regression, with the respective codes as dependent variables and countries as independent variables. Finally, we assess to what extent the substantive codes can explain the variance of the answers for the "civil disobedience" item. For that purpose, we run separate regression analyses for the category-selection probe and the comprehension probe. The substantive codes serve as independent variables in these analyses; the "civil disobedience" item is the dependent variable. Further socio-demographic items serve as control variables.

3 Results

3.1 Comparison of ISSP and Web survey data

To begin with, we compare the ISSP and the Web survey results on the "rights of people in a democracy" scale (Table 1). The comparison shows that the striking patterns found in the ISSP are neatly reproduced in the Web surveys. That is, the value for the index (items a–e) is about the same for all countries/regions. The right to engage in civil disobedience is supported less than the other rights in all the countries. The biggest gap between the index and the civil disobedience mean is once again found in Canada, Denmark, and the U.S. Furthermore, similar to the ISSP, the "civil disobedience" item is marked by a high non-response rate which is overall 7.4 %. For the other items of the "rights in a democracy" scale, non-response reaches maximally 2.2 % in a country. In fact, many respondents indicated that they did not

understand the question and/or the term "civil disobedience" and, therefore, were not able to provide an answer.

3.2 Coding scheme for the category-selection probe

Next, we present the substantive coding scheme for the category-selection probe (see Table 3, which also lists examples for each of the codes). The codes for the category-selection probe stand for reasons—in the widest sense—that respondents put forward to explain their answer value regarding civil disobedience. In line with our assumptions that a low level of trust in government plays a role in the civil disobedience rating, we were able to identify answers that went into this direction. We distinguished two dimensions. The code NO RESPONSIVENESS assembles answers that focus on problems with vertical accountability between politicians and respondents, on the lack of politicians' responsiveness towards voter's needs and on serving only the interests of big business instead of those of the voters. The code GENERAL DISSATISFACTION represents discontent with the government in a more general way without respondents accusing the government to act against voter's needs, breaking election promises or serving only big business. The focus with GENERAL DISSATISFACTION is more on the importance of having a method at one's disposal that helps to show dissatisfaction—whenever a situation of discontent with the government arises. Admittedly, it was not always easy to draw a line between these two codes. A clear-cut example for NO RESPONSIVENESS would be (answer edited for spelling and punctuation mistakes): "If you look at 'empires' of the past, you will see that corruption from within was their downfall. Now look at how things are today. Government officials abuse their power for their own gain. They are too busy satisfying their own needs, the needs of special interest, and no longer care for the working man. Even though they are the backbone of this nation." (U.S., 2868). A clear-cut example for GENERAL DISSATISFACTION would be: "People must have a way of showing governnment when the latter are making the wrong decisions" (Canada, 5703). A not that straightforward case (coded as GENERAL DISSATISFACTION) was: "You should be allowed to protest when the government does not put the public first" (Canada, 9626). Inter-rater agreement helped to assess and improve our coding in the critical cases. In addition, we ran analyses also for both dimensions merged to ensure that critical coding cases did not introduce measurement artifacts.

Another fundamental line of reasoning was referring to freedom of speech, to one's right to protest, one's right to participate in decision-taking, etc. (RIGHTS). Furthermore, we had respondents defining or describing what civil disobedience should (not) look like and under which conditions it would be (un-)acceptable (RANGE OF ACCEPTABILITY). Another group of respondents preferred other means to be taken (first) in order to put their message across, such as elections or available legal routes (OTHER METHODS). Still another group referred to respecting and accepting authority and the elected government, taking into account that not all activities please everyone (ACCEPTANCE). Also, some respondents explicitly mentioned violence or destruction that go along with civil disobedience (EXPLICIT VIOLENCE). Respondents also more implicitly referred to violence and destruction in the sense that they were aware that the actions do not always end on a peaceful note and that they took this into account when answering (IMPLICIT VIOLENCE). Then, there were respondents who listed negative effects more in general, effects other than explicit violence and/or destruction (NEGATIVE EFFECTS), for example, disruption or disturbance of the public or the political system in general. Some respondents argued that civil disobedience has no effect whatsoever or is of no use (NO EFFECT), while others again spoke of positive effects of civil disobedience (POSITIVE EFFECTS). Finally, we had a group of respondents who

Table 3 Substantive coding scheme for the category-selection probe

	Category selection probe	Example
NO RESPONSIVENESS	Problems with vertical accountability/responsiveness of politicians	The government like to disregard public opinions and do as they please. When it comes down to it, the public has to resort to such acts as to show the government that they should have a say in these matters. (CA, 5410) Governments don't have a carte blanche to do whatever it is they want to do. Governments and politicians have long since forgotten they are here to represent the citizens, all they are interested in now is money and power. (CA, 4934)
GENERAL DISSATISFACTION	Dissatisfaction with government in general	Unjust laws should be opposed and changed. (CA, 5426) Puts pressure on the government to make a decision and also to show them that the people are restless. (CA, 4995) Sometimes the government needs to take notice. (CA, 5743)
RIGHTS	Freedom of speech/right to protest/right to voice one's opinion	People should have the right to have their points of view listened to and heard. (CA, 5186) It's freedom of speech. (CA, 5062)
RANGE OF ACCEPTABILITY	Ideas on what civil disobedience should (not) look like	Depends whether they are doing so in a peaceful manner or not...I agree completely with peaceful protest. (CA, 5905) The people need to express their views but need to do so within the law. (CA, 5474)
OTHER METHODS	Other means better/first to be used	They should not. They have the opportunity to bring in new people at elect time. (CA, 4295) Civil disobedience is never good, peaceful protest is a better way. (CA, 5825)
ACCEPTANCE	Acceptance of government/authority	I think that if it is a government action that it is basically a final statement. (CA, 5600)
EXPLICIT VIOLENCE	Negative effects—violence/destruction explicit	There are a lot more ways to oppose a government. Civil disobedience can quickly slip into rioting, looting and injuries. (CA, 5111)
IMPLICIT VIOLENCE	Negative effects—violence/destruction implicit	It's important, as long as the citizens don't harm others or property. (CA, 5191)
NEGATIVE EFFECTS	Negative effects—other	There is no reasons in the world that humans should partake in civil disobedience.. it just makes things worse.. such as curfews, loss of freedom of speech.. etc... (CA, 6123) People should not partake in this kind of act. Leads to more confusion and more trouble (CA, 7526)

Table 3 continued

	Category selection probe	Example
NO EFFECT	No effects/no use	Acts of civil disobedience have never gotten people anywhere. (CA, 4953)
POSITIVE EFFECT	Positive effects	Historically, progressive change happens when the people vocally challenge authority, and fight for their rights. (CA, 5656)
AMBIVALENCE	Ambivalence of term 'civil disobedience'	This was selected because the definition of civil disobedience is not clear, therefore a definite number could not be chosen. Some people feel that it means being violent which should not be tolerated. (CA, 5433)
OTHER	Other	There should never be civil disobedience other than under a ruthless dictatorship or puppet government. (CA, 5237)

Note: Examples are partly assigned to more codes than the code under which they are listed. Examples are taken from Canada (spelling and punctuation were corrected)

emphasized the ambiguous or unclear meaning of the term civil disobedience (AMBIVALENCE). All other answers which could not be coded into any of the above codes were assigned to the OTHER code. In general, answers were either assigned to non-response codes or to (a combination of) the substantive codes. AMBIVALENCE and OTHER, however, could not be combined with any other code.

3.3 Coding scheme for the comprehension code

The codes for the comprehension probe represent what respondents associate with the term "civil disobedience". The codes as well as examples are presented in Table 4. In line with our assumptions about negative associations, we were able to discern the code VIOLENCE/DESTRUCTION where respondents name violent and/or destructive activities, such as rioting or looting. Similarly, we were able to identify a code that addresses disturbances more in general, such as blocking streets or other disruptions (DISTURBANCE). The counter-code to VIOLENCE/DESTRUCTION assembles answers where respondents explicitly refer to the peacefulness of activities (PEACEFUL), such as "nonviolent protest". We had an additional code for respondents who listed activities ranging from peaceful to violent (PEACEFUL/VIOLENCE). A large group of respondents simply listed activities *without* making any explicit reference to peacefulness, violence/destruction or disturbances, such as "demonstrating, picketing, writing letters" or "sit-in, marches" (LISTING ACTIVITIES). Demonstrations or marches are usually peaceful events, but without any explicit reference to peacefulness we did not assign such answers to the code PEACEFUL. Equally, without explicitly referring to disturbances we did not assign such answers to the code DISTURBANCE—in practice, of course, each demonstration or sit-in will somehow be a disturbance to a certain group of people. Then, there were respondents who thought of breaking the law (BREAKING LAW) and those who thought of breaking rules more in general (BREAKING

Table 4 Substantive coding scheme for the comprehension probe

	Comprehension probe	Example
VIOLENCE/DESTRUCTION	Explicitly referring to violence/destruction	Protest that ends or conveys violence. (U.S., 2729) Looting, anarchy, rebellion. (U.S., 2764) Violence and derogatory statements or words. (U.S., 2869)
DISTURBANCE	Explicitly referring to disturbances other than violence/destruction	Blocking street's and blocking doors to courts. (U.S., 6091) Civil disobedience, to me, means taking an active role in disrupting societally established standards, such as shutting down public transportation or causing other disruptions that directly affect persons, usually a majority, who disagree with their positions. (U.S., 3620)
PEACEFUL	Explicitly referring to peaceful activities/to the absence of violence	Non-violent protests (U.S., 2714) Actions of the 1960s like civil rights and anti-war protests that were entirely non violent are examples of acceptable 'civil disobedience'. Any action that involves violence of any kind is unacceptable and must be treated as such. (U.S., 2865)
PEACEFUL/VIOLENCE	Activities ranging from peaceful to destructive/violent	Riots, picketing, violence, public displays/demonstrations. (U.S., 2713) Protesting and rioting. (U.S., 2817)
LISTING ACTIVITIES	Mere listing of activities	Picketing, forums, boycotts, petitions, marches. (U.S., 2906) Protesting, writing letters, not taking part. (U.S., 2987) Sit ins, marches. (U.S., 4243)
BREAKING LAW	Breaking the law	Refusal to obey civil laws that one views as unjust in a passive manner. Example would be public protests. (U.S., 2874) I see civil disobedience as rioting, looting, burning the flag, committing illegal acts, acting in ways that harm others or show disregard for the law. (U.S., 8397) Disobeying the law. (U.S., 8274)

Table 4 continued

	Comprehension probe	Example
BREAKING RULES	Breaking rules	Breaking rules. (U.S., 2775)
COPY-PASTE	Copy-paste from the Internet	Civil disobedience is the active, professed refusal to obey certain laws, demands, and commands of a government, or of an occupying international power. (CA, 5387)
OTHER	Other	Civil rights movement, gay rights, abortion rights, (U.S., 2941) If citizens don't get a chance to speak up on opposed government acts how is it any different than a dictatorship. (U.S., 4296)

Note: Examples are partly assigned to more codes than the code under which they are listed. Examples are taken from Canada and the U.S. (spelling and punctuation were corrected)

RULES). These categories could not always be strictly separated on the semantic level. Also here, we merged these codes for additional analyses to make sure that our differentiation did not introduce any bias. With the comprehension probe we also had respondents displaying general or more specific dissatisfaction with the government. However, since it was not always clear to us whether respondents' dissatisfaction was part of a definition they gave us for the term "civil disobedience" or whether they merely wanted to expose their opinion (e.g., "I just believe that people elect the politicians, so if they do not do the job they said they would then we as a people have the right to kick them out of office and find someone new to do their job right." U.S., 11499), and since reliable coding proved difficult in some of these cases, we eventually abstained from coding these dissatisfaction answers separately and used the OTHER code instead. Furthermore, we grouped answers together that were—demonstrably—copied from the Web (COPY-PASTE). However, this code let to only 24 answers out of more than 1,800 being the result of copy-paste activities. The code OTHER rounded up the coding scheme: answers that did not fall into the above listed codes and answers that particularly belonged to the realm of an opinion statement were assigned to this code. In general, answers were either assigned to non-response codes or (a combination of) the substantive codes. COPY-PASTE and OTHER were single-coded and distinctive codes.

3.4 Distribution of substantive codes for the category-selection probe

Table 5 shows the distribution of codes for the category-selection probe. The columns do not add up to 100 % since multiple coding was allowed.

The country pattern for NO RESPONSIVENESS (first line in Table) is most revealing. This code not only has a high (overall) frequency but it also mirrors the divide between the two groups of countries that we have already seen in the data of the closed item. While only between 4.2 and 6.5 % of American, Danish, and Canadian respondents express their discontent over a lack of vertical accountability or responsiveness, between 11.6 and 18.0 % of German, Hungarian, and Spanish respondents do so. A logistic regression with the code NO RESPONSIVENESS as dependent variable and country as independent variable demonstrates that these country differences are significant (model not presented). The code GENERAL DISSATISFACTION has the highest (overall) prevalence. Also here, we see a

Table 5 Distributions of the category-selection probe answers across countries in % (actual numbers in parentheses)

Code	Canada	Denmark	Germany (eastern)	Germany (western)	Hungary	Spain	U.S.	Total
NO RESPONSIVENESS	4.2 (10)	6.5 (15)	18.0 (47)	14.8 (33)	14.8 (41)	11.6 (29)	4.2 (10)	10.8 (185)
GENERAL DISSATISFACTION	14.6 (35)	17.0 (39)	19.9 (52)	16.6 (37)	18.1 (50)	19.6 (49)	13.7 (33)	17.1 (295)
RIGHTS	9.6 (23)	9.1 (21)	8.8 (23)	13.0 (29)	13.4 (37)	6.4 (16)	10.0 (24)	10.1 (173)
RANGE OF ACCEPTABILITY	16.7 (40)	12.2 (28)	4.2 (11)	7.6 (17)	5.4 (15)	10.8 (27)	12.9 (31)	9.8 (169)
OTHER METHODS	7.1 (17)	6.1 (14)	3.1 (8)	1.8 (4)	1.1 (3)	4.8(12)	5.4 (13)	4.1 (71)
ACCEPTANCE	1.3 (3)	5.2 (12)	3.8 (10)	4.0 (9)	2.9 (8)	1.2 (3)	2.1 (5)	2.9 (50)
EXPLICIT VIOLENCE	6.3 (15)	2.2 (5)	0.0 (0)	3.1 (7)	1.1 (3)	1.6 (4)	3.3 (8)	2.4 (42)
IMPLICIT VIOLENCE	13.4 (32)	6.1 (14)	2.7 (7)	3.1 (7)	2.5 (7)	5.6 (14)	13.3 (32)	6.6 (113)
NEGATIVE EFFECTS	5.0 (12)	3.5 (8)	2.3 (6)	4.0 (9)	4.0 (11)	3.6 (9)	5.8 (14)	4.0 (69)
NO EFFECT	3.4 (8)	0.0 (0)	0.8 (2)	0.0 (0)	1.1 (3)	0.8 (2)	4.2 (10)	1.5 (25)
POSITIVE EFFECT	1.7 (4)	0.0 (0)	0.8 (2)	1.8 (4)	1.1 (3)	0.4(1)	1.7 (4)	1.1 (18)
AMBIVALENCE	3.4 (8)	6.1 (14)	3.1 (8)	2.2 (5)	1.4 (4)	0.8 (2)	4.6 (11)	3.0 (52)
OTHER	10.9 (26)	16.1 (37)	13.8 (36)	17.0 (38)	20.2 (56)	16.0 (40)	12.5 (30)	15.3 (263)
Probe non-response	28.9 (69)	26.5 (61)	28.7 (75)	26.0 (58)	20.9 (58)	30.0 (75)	31.1 (75)	27.4 (471)

Note: n = 1,721, respondents having values 1–7 on closed civil disobedience item, excl. DKs. The basis for each percentage is the number of respondents in the category-selection split in each country. The entries in the rows do not add up to 100 % since multiple coding was allowed

tendency in line with the peculiar pattern, but the country differences are not significant in this case. Furthermore, the codes RANGE OF ACCEPTABILITY and IMPLICIT VIOLENCE show a striking pattern (country differences significant for both codes). Only between 4.2 and 7.6 % name a RANGE OF ACCEPTABILITY in Germany and Hungary (e.g., "only within the law", HU, 3197) or "all needs to remain within an appropriate scope—violence is never a solution", DE, 6171). In the other countries, this code has an occurrence of 10.8 to 16.7 %. IMPLICIT VIOLENCE predominates in Canada and the U.S. with around 13 % compared to 6 % and lower in the other countries. The high levels of both RANGE OF ACCEPTABILITY and IMPLICIT VIOLENCE for the U.S. and Canada can be explained by the coding rules. Whenever the range of acceptability was defined in terms of peacefulness, both codes were given; e.g., "Citizens should be allowed to engage in acts of civil disobedience up to a point, but not go as far as things becoming violent" (U.S., 2898). Finally, the code OTHER METHODS was hardly mentioned in Germany and Hungary compared to the other countries (significant country differences).

3.5 Distribution of substantive codes for the comprehension probe

Table 6 shows the distribution of codes for the comprehension probe. Once again, the columns do not add up to 100 % since multiple coding was allowed.

The most remarkable code for the comprehension code is VIOLENCE/DESTRUCTION. Canada and the U.S. in particular are set apart from the others. Respondents from Canada and the U.S. name destructive/violent activities much more often than respondents in the other countries. Logistic regression with VIOLENCE/DESTRUCTION as dependent variable reveals that the country difference is significant (model not presented). When it comes to explicitly mentioning the peacefulness of civil disobedience or the peacefulness/violence range (PEACEFUL & PEACEFUL/VIOLENCE), Canada and the U.S. are again on top (significant country differences). This only highlights the importance that the dimension of peacefulness vs. violence has for Americans and Canadians when it comes to civil disobedience. BREAKING LAW particularly dominates in Denmark and Spain with around 24 % compared to 8–16 % in the other countries (significant country differences).

3.5.1 Regression of civil disobedience items on substantive codes

The substantive codes were subsequently used in regression analyses. We start with the category-selection split (Table 7).

In model 1, we regressed the civil disobedience item on the countries. Significant differences were found between the U.S. (the baseline) and Spain, Hungary, and (western/eastern) Germany. Denmark and Canada did not differ from the U.S. The explained variance is 12 percent. In model 2, we added all the substantive category-selection codes. The country differences were markedly reduced but remained significant. The explained variance rose to 36 %. Since the codes alone explain almost 30 % of variance, the second model leaves us with almost 6 % of variance that is due to the remaining country effect. Codes in model 2 which had significant effects on civil disobedience and, at the same time, large enough case numbers which also differed across countries (see Table 5 on country distributions) were: NO RESPONSIVENESS, GENERAL DISSATISFACTION, RIGHTS, and OTHER METHODS. The first three codes increased the support for civil disobedience, while with the forth code the support for civil disobedience dwindled. In model 3, we added socio-demographic variables (sex, education, age, but also left-right orientation and political interest) as control variables. While a political tendency towards the right significantly decreased the score on

Table 6 Distributions of the comprehension probe answers across countries in percent (actual numbers in parentheses)

Code	Canada	Denmark	Germany (eastern)	Germany (western)	Hungary	Spain	U.S.	Total
VIOLENCE/DESTRUCTION	24.9 (60)	7.0 (17)	5.4 (13)	2.5 (6)	2.5 (6)	7.7 (19)	21.4 (55)	10.3 (176)
DISTURBANCE	4.2 (10)	6.2 (15)	1.7 (4)	2.1 (5)	4.6 (11)	1.6 (4)	3.9 (10)	3.5 (59)
PEACEFUL	11.6 (28)	3.7 (9)	3.8 (9)	8.4 (20)	6.3 (15)	4.1 (10)	17.1 (44)	7.9 (135)
LISTING ACTIVITIES	23.7 (57)	32.1 (78)	32.2 (77)	36.3 (86)	42.3 (101)	22.4 (55)	21.4 (55)	29.9 (509)
PEACEFUL/VIOLENCE	9.5 (23)	1.7 (4)	2.1 (5)	0.4 (1)	0.0 (0)	0.8 (2)	6.2 (16)	3.0 (51)
BREAKING LAW	7.5 (18)	24.3 (59)	9.2 (22)	8.4 (20)	15.5 (37)	23.6 (58)	10.5 (27)	14.2 (241)
BREAKING RULES	0.8 (2)	4.5 (11)	2.9 (7)	4.2 (10)	3.4 (8)	4.5 (11)	0.4 (1)	2.9 (50)
COPY-PASTE*	1.7 (4)	1.2 (3)	1.7(4)	0.0 (0)	1.3 (3)	1.2 (3)	0.0 (0)	1.0 (17)
OTHER	13.7 (33)	11.1 (27)	20.9 (50)	17.7 (42)	16.7 (40)	18.7 (46)	13.2 (34)	16.0 (272)
Probe non-response	20.3 (49)	20.2 (49)	27.6 (66)	26.6 (63)	18.8 (45)	24.4 (60)	20.6 (53)	22.6 (385)

Note: n = 1,702, respondents having values 1–7 on closed civil disobedience item, excl. DKs. COPY-PASTE*: according to manual coding; few further culprits were identified by keystroke logging (total identified so far: 24). The basis for each percentage is the number of respondents in the comprehension split in each country. The entries in the rows do not add up to 100 % since multiple coding was allowed

Table 7 Linear regression analysis for countries and argumentation patterns of the category-selection probe predicting attitudes on civil disobedience (unstandardized coefficients, standard error in parentheses)

	Model 1 countries only		Model 2 countries and codes		Model 3 countries and codes and socio-demographics	
Countries						
Base line: U.S						
Canada	−0.15	(0.16)	−0.18	(0.13)	−0.19	(0.13)
Denmark	0.08	(0.16)	−0.03	(0.13)	−0.05	(0.13)
Germany (eastern)	1.06***	(0.16)	0.67***	(0.14)	0.64***	(0.14)
Germany (western)	1.08***	(0.15)	0.67***	(0.13)	0.62***	(0.13)
Hungary	1.49***	(0.15)	1.06***	(0.13)	1.07***	(0.13)
Spain	0.69***	(0.15)	0.45***	(0.13)	0.42**	(0.13)
Argumentation patterns						
No responsiveness			1.64***	(0.12)	1.60***	(0.12)
Gen. dissatisfaction			1.29***	(0.10)	1.27***	(0.10)
Rights			1.31***	(0.12)	1.30***	(0.12)
Range of acceptability			−0.04	(0.14)	−0.04	(0.15)
Other methods			−1.34***	(0.19)	−1.35***	(0.19)
Acceptance			−1.15***	(0.21)	−1.16***	(0.21)
Explicit violence			−0.28	(0.23)	−0.25	(0.23)
Implicit violence			−0.20	(0.17)	−0.22	(0.17)
Negative effects			−0.97***	(0.18)	−0.98***	(0.18)
No effect			−1.08***	(0.29)	−1.12***	(0.29)
Positive effects			0.82*	(0.34)	0.77*	(0.34)
Ambivalence			0.03	(0.21)	0.02	(0.21)
Other			0.44***	(0.11)	0.42***	(0.11)
Control variables						
Left–right					−0.04**	(0.02)
Political interest					0.01	(0.03)
Sex (base line: men)					0.02	(0.07)
Education (base line: lower education)					−0.06	(0.07)
Age (centered)					0.003	(0.002)
_cons	4.46***	(0.11)	4.25***	(0.11)	4.50***	(0.20)
N	1,598		1,598		1,598	
R^2	0.120		0.362		0.366	
adj. R^2	0.116		0.355		0.357	

the civil disobedience item, the other variables did not impact on civil disobedience. The increase in explained variance in model 3 is practically zero. More importantly, the overall results produced in model 2 did not change with the introduction of the control variables.

We ran additional analyses with, on the one hand, codes GENERAL DISSATISFACTION and NO RESPONSIVENESS merged and, on the other hand, codes GENERAL DISSAT-

Table 8 Regression analysis for countries and argumentation patterns of the comprehension probe predicting attitudes on civil disobedience (unstandardized coefficients, standard error in parentheses)

	Model 1 countries only		Model 2 countries and codes		Model 3 countries and codes and socio-demographics	
Countries base line: U.S.						
Canada	−0.15	(0.15)	−0.08	(0.14)	−0.09	(0.14)
Denmark	−0.10	(0.15)	−0.15	(0.14)	−0.17	(0.14)
Germany (eastern)	0.98***	(0.15)	0.85***	(0.14)	0.78***	(0.14)
Germany (western)	1.07***	(0.15)	0.87***	(0.14)	0.79***	(0.14)
Hungary	1.32***	(0.15)	1.12***	(0.14)	1.15***	(0.14)
Spain	0.62***	(0.15)	0.62***	(0.14)	0.56***	(0.14)
Argumentation patterns						
Violence/destruction			−1.55***	(0.16)	−1.56***	(0.16)
Disturbance			0.12	(0.20)	0.08	(0.20)
Peaceful			0.98***	(0.15)	0.90***	(0.15)
Listing activities			0.51***	(0.10)	0.46***	(0.10)
Peaceful/violence			1.09***	(0.26)	1.09***	(0.26)
Breaking law			−0.10	(0.12)	−0.12	(0.12)
Breaking rules			0.08	(0.22)	0.07	(0.22)
Copy-paste			0.35	(0.41)	0.31	(0.40)
Other			0.37**	(0.12)	0.34**	(0.12)
Control variables						
Left–right					−0.07***	(0.02)
Political interest					0.13***	(0.04)
Sex (base line: men)					−0.02	(0.08)
Education (base line: lower education)					−0.18*	(0.08)
Age (centered)					−0.002	(0.003)
_cons	4.51***	(0.10)	4.42***	(0.12)	4.55***	(0.21)
N	1,591		1,591		1,591	
R^2	0.113		0.226		0.242	
adj. R^2	0.110		0.219		0.232	

ISFACTION, NO RESPONSIVENESS, and RIGHTS merged to control for any potential bias that might have arisen through coding. The effects described above could neatly be replicated. Additionally, we ran the analyses also with a binary non-response variable (non-response for the probe) added as a predictor. There are practically no changes in explained variance compared to the models above, but we find that those who failed to provide a substantive response to the probe have in general more positive attitudes with regard to civil disobedience (0.42; $p < 0.01$).

Similar analyses for the comprehension answers led to the following results (see Table 8): Introducing country as independent variable in model 1 led to 11 % of explained variance for civil disobedience. Upon adding the comprehension code answers to the regression, the

country differences could again be reduced but to a smaller degree than was the case for the category-selection probe split. The explained variance in model 2 rose to 22 %. With 14 % of explained variance through the substantive codes alone, this leaves us, in model 2, with 8 % of variation due to the remaining country effect. The following codes were found to have a significant impact on civil disobedience and were at the same time of frequent occurrence across countries: VIOLENCE/DESTRUCTION, PEACEFUL, LISTING ACTIVTIES, and PEACE/VIOLENCE. While with the first of these codes the support for civil disobedience decreased, with the other three codes the support went up. Upon introducing the socio-demographic variables in model 3, we find that placing oneself towards the political right and higher education decrease the civil disobedience value. On the contrary, a higher political interest goes hand in hand with increased importance rating for civil disobedience. Overall, however, the socio-demographic variables do not change the patterns found in the previous models, and the explained variance only rises by 1 % in model 3.

We also set up regression models with probe non-response as additional independent variable besides the substantive codes. Those who failed to give a substantive answer to the comprehension probe did not differ significantly from the other respondents. Apart from that, results from the above models could be replicated. The replication of the above results was also possible when merging the codes BREAKING LAW and BREAKING RULES and introducing them into regression analyses.

In sum, both coding schemes—for the category-selection and the comprehension probe—helped to reduce the pure country effects, the category-selection answers more so than the comprehension answers.

4 Discussion

We set out to shed light on interpretation differences across countries for the civil disobedience item of the ISSP by making use of Web probing. Precondition for any meaningful use of the Web survey data to explain the ISSP data was the replication of the country-specific answer pattern for both the "rights in a democracy" scale and the "civil disobedience" item. We were successful in achieving this. The first step towards answering our research questions was then to code the answers to the category-selection and the comprehension probe.

The regression analyses with the category-selection probe answers revealed that a low level of trust in government, notably expressed through the codes NO RESPONSIVENESS and GENERAL DISSATISFACTION, were the most obvious drivers for the high importance rating of the civil disobedience item. The country-specific distributions for these codes were in line with the peculiar civil disobedience pattern (U.S., Canada, and Denmark vs. Spain, (eastern/eastern) Germany, and Hungary). We, therefore, suggest that they explain part of the country differences for the closed item. In particular, the lack of responsiveness of politicians towards their electorate in Germany, Hungary and, to a weaker degree, in Spain seems to have played a decisive role in the civil disobedience rating. The differences in the level of trust are essentially substantive differences which reflect—save for imprecise measurement—real country differences. They should not endanger the equivalence status of the item, that is, they should not put into question the comparability of the item.

On the contrary, the regression analyses with the comprehension probe answers brought to light a critical equivalence issue. Interpretation patterns involving a violence/peace divide were the strongest predictors for the civil disobedience score. Especially the code VIOLENCE/DESTRUCTION can be named in this context. Canadians and U.S.-Americans associate civil disobedience much more frequently with violence and destruction than respon-

dents in the other countries. We suggest that these negative associations led in particular to the low importance rating of civil disobedience in Canada and the U.S. With many respondents in certain countries thinking of destructive and violent actions while this does not or hardly happen with respondents in other countries, we are faced with different understandings of the term "civil disobedience" (that is, we essentially have a different attitude object) and, thus, with the lack of equivalence.

These findings in combination bring us to the conclusion that the low civil disobedience values for Canada and the U.S. seem to result from both valid country differences (somewhat better perception of politicians in Canada and the U.S. than in the other countries) and a methods artifact (associations of violence/destruction that systematically add an interpretation shade that is hardly existing in the other countries). For Denmark, the low value might be due to mainly peaceful associations with civil disobedience, coupled with higher satisfaction with politicians than in Germany, Hungary or in Spain. Any future use of the item should specify the form of civil disobedience to ensure similar understanding and comparability across countries.

4.1 Limitations

In our study, the category-selection probe answers were more successful in reducing country effects than the comprehension probe answers. However, since we implemented the two probes in different splits and thus only had respondents answering either of the probes, we do not know how argumentation and interpretation interact within subjects. The different patterns could be additive but also overlapping. Future studies may combine the two probes in one person in order to allow the assessment of the relative explanatory power of the diverse patterns.

The Web probing method itself has some limitations. Although the majority of answers helped us to uncover country-specific interpretation patterns and thus served our needs, we need to issue a cautionary note. Respondents do not necessarily give answers that match the type of probing, not to mention non-response. A comprehension probe may thus lead, for instance, to elaborations on respondents' opinions rather than to respondents' definitions of certain terms. Such deviating behavior requires further investigation in the future.

Related to the issue above is the fact that respondents can search the Web for definitions of terms when asked a comprehension probe. Although this did not seem problematic in our study, it has to be considered when implementing comprehension probes in Web studies.

As with all probing techniques, one cannot fully be sure that the definitions or rationalizations given by the respondents are indeed what was going on in respondents' minds when answering the closed-ended survey questions. Still, the major answer patterns presented in this study (i.e., NO RESPONSIVENESS and GENERAL DISSATISFACTION for the category-selection probe, and VIOLENCE/DESTRUCTION for the comprehension probe) seem to be a plausible explanation for the answer values chosen by the respondents.

Our analytical approach itself, i.e., to offer answers on what might have happened in the ISSP based on conclusions from Web survey data, includes a mode switch, the use of both a representative and a non-representative survey, and a time lag. Our conclusions thus have to be treated with caution, although the replication of the two-group pattern for the civil disobedience item speak in favor of our approach. We recommend that in the future the time lag between a representative survey and a Web survey should be reduced. Alternative approaches may include follow-up Web surveys with the same respondents from a representative field or employing a probability-based Web panel, if available.

Talking about the future application of the method: Web probing, rather than being implemented after a survey, could equally be used as a pure pretesting tool and thus with a view to evaluate questionnaires and modify them for future use. The time needed to analyze (and code) hundreds of answers across countries, including the time potentially needed for translation, may be more than usually available during the questionnaire design phase, though. The method could equally be used to assess (changes in) item interpretation across countries at regular intervals.

Despite our general optimism with the method, we wish to stress that Web probing cannot replace traditional face-to-face cognitive interviewing when in particular in-depth information on response processes is sought that can only be obtained with several follow-up probes, when emergent probes (developed flexibly based on the behavior of the subject, Willis 2005) are considered the prevailing cognitive interviewing paradigm or when particular groups are targeted that cannot be reached via Web surveys. Still, cross-national Web probing has its potential and deserves further attention in practice and research.

Acknowledgments This work was supported by the German Research Foundation (DFG) as part of the PPSM Priority Programme on Survey Methodology (SPP 1292) [project # 574187].

References

Beatty, P.C., Willis, G.B.: Research synthesis: the practice of cognitive interviewing. Public Opin. Quart. **71**, 287–311 (2007)

Behr, D., Kaczmirek, L., Bandilla, W., Braun, M.: Asking probing questions in web surveys: which factors have an impact on the quality of responses? Soc. Sci. Comput. Rev. (2012). doi:10.1177/0894439311435305

Behr, D., Braun, M., Kaczmirek, L., Bandilla, W.: Testing the validity of gender ideology items by implementing probing questions in web surveys. Field Methods (Forthcoming)

Bolzendahl, C., Coffe, H.: A gender gap in citizenship norms? The importance of political, civil and social rights and responsibilities. Br. J. Sociol. **60**(4), 763–791. http://escholarship.org/uc/item/71x598x2 (2008). Accessed 24 April 2012

Childs, J., Goerman, P.: Bilingual questionnaire evaluation and development through mixed pretesting methods: the case of the U.S. census nonresponse followup instrument. J. Off. Stat. **26**, 535–557 (2010)

DeMaio, T.J., Rothgeb, J.M.: Cognitive interviewing techniques: in the lab and in the field. In: Schwarz, N., Sudman, S. (eds.) Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research, pp. 177–195. Jossey-Bass, San Francisco (1996)

Donovan, T., Denemark, D., Bowler, S.: Trust in government. The United States in comparative perspective. http://faculty.wwu.edu/donovat/trust.pdf (2008). Accessed 24 April 2012

Fitzgerald, R., Widdop, S., Gray, M., Collins, D.: Identifying sources of error in cross-national questionnaires: application of an error source typology to cognitive interview data. J. Off. Stat. **27**, 569–599 (2011)

Forsyth, B.H., Stapleton Kudela, M., Levin, K., Lawrence, D., Willis, G.B.: Methods for translating an English-language survey questionnaire on tobacco use into Mandarin, Cantonese, Korean, and Vietnamese. Field Methods **19**, 264–283 (2007)

Goerman, P.L., Caspar, R.A.: A preferred approach for the cognitive testing of translated materials: testing the source version as a basis for comparison. Int. J. Soc. Res. Methodol. **13**, 303–316 (2010)

Harkness, J.A., Edwards, B., Hansen, S.E., Miller, D.R., Villar, A.: Designing questionnaires for multipopulation research. In: Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., Smith, T.W. (eds.) Survey Methods in Multinational, Multiregional, and Multicultural Contexts., pp. 33–57. Wiley, Hoboken (2010)

ISSP Research Group: International social survey programme 2004: citizenship (ISSP 2004). Source questionnaire. GESIS Data Archive, Cologne (2004a)

ISSP Research Group: International social survey programme 2004: citizenship (ISSP 2004). Data file Vers. 1.2.0. GESIS Data Archive, Cologne. doi:10.4232/1.10078 (2004b)

Lee, J.: Conducting cognitive interviews in cross-national settings. Assessment (2012). doi:10.1177/1073191112436671

Miller, K., Fitzgerald, R., Padilla, J.-L., Willson, S., Widdop, S., Caspar, R., Dimov, M., Gray, M., Nunes, C., Prüfer, P., Schöbi, N., Schoua-Glusberg, A.: Design and analysis of cognitive interviews for comparative multinational testing. Field Methods **23**, 379–396 (2011)

Morren, M., Gelissen, J., Vermunt, J. K.: Exploring the response process of culturally differing survey respondents with a response style: a sequential mixed-methods study. Field Methods (Forthcoming)

Prüfer, P., Rexroth, M.: Kognitive Interviews [cognitive interviews]. ZUMA How-to-Reihe 15. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf?download=true (2005). Accessed 24 April 2012

Reeve, B.B., Willis, G., Shariff-Marco, S.N., Breen, N., Williams, D.R., Gee, G.C., Alegría, M., Takeuchi, D.T., Stapleton, M., Levin, K.Y.: Comparing cognitive interviewing and psychometric methods to evaluate a racial/ethnic discrimination scale. Field Methods **34**, 397–419 (2011)

Schuman, H.: The random probe: a technique for evaluating the validity of closed questions. Am. Sociol. Rev. **31**, 218–222 (1966)

Schwarz, N.: Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation. Erlbaum, Mahwah (1996)

Strack, F., Martin, L.: Thinking, judging, and communicating: a process account of context effects in attitude surveys. In: Hippler, H.-J., Schwarz, N., Sudman, S. (eds.) Social Information Processing and Survey Methodology, pp. 123–148. Springer, New York (1987)

Thrasher, J.F., Quah, A.C.K., Dominick, G., Borland, R., Driezen, P., Awang, R., Omar, M., Hosking, W., Sirirassamee, B., Boado, M.: Using cognitive interviewing and behavioral coding to determine measurement equivalence across linguistic and cultural groups: an example from the international tobacco control policy evaluation project. Field Methods **23**, 439–460 (2011)

Tourangeau, R., Rips, L.J., Rasinski, K.: The Psychology of Survey Response. Cambridge University Press, Cambridge (2000)

Willis, G.B.: Cognitive Interviewing: A Tool for Improving Questionnaire Design. Sage, Thousand Oaks (2005)

Willis, G.B., Lawrence, D., Hartman, A., Stapleton Kudela, M., Levin, K., Forsyth, B.: Translation of a tobacco survey into Spanish and Asian languages: the tobacco use supplement to the current population survey. Nicot. Tob. Res. **10**, 1075–1084 (2008)