2

Constructing the Questionnaire:

The Challenges of Measuring Attitudes Towards Democracy Across Europe

Lizzy Winstone, Sally Widdop, and Rory Fitzgerald

NB – final draft not published version as in press at time of submission

INTRODUCTION

This chapter explores the challenges of designing questions to measure attitudes towards democracy in a cross-national survey. The European Social Survey (ESS) has often included topics that are not generally part of the day-to-day discourse of many respondents, such as ageism or economic morality. However, in many ways a module focused on respondents' understandings and evaluations of democracy was particularly challenging to include since the detail of the topic was unlikely to be something that many potential respondents would have considered in detail. This chapter addresses the specific challenges of designing questions measuring attitudes to democracy, in particular decisions about the concepts to include or exclude, social desirability, and response formats. This chapter explores the decision-making during the design process, the need to strike a balance between theoretical measurement aims and what is practical to measure in a general social survey, as well as the attempt to strike a balance between different forms of measurement error.

The questionnaire module "Europeans' understandings and evaluations of democracy" was included as one of two rotating modules in Round 6 of the ESS, which was fielded in most of the participating twenty-nine countries between September and December 2012 (for details of the preparation of this round, see Appendix A). The ESS Round 6 democracy module centers around nineteen core concepts referring to different features of democracy, which have been briefly introduced in the introduction to this book and which shall be described in detail in Chapter 3, as well as a broader concept, *support for democracy*, including questions on the overall importance of living in a democracy and the overall satisfaction with democracy (SWD) in respondents' countries. The module systematically distinguishes between items addressing the respondents' views of democracy, and items addressing the way they evaluate the democracy in their own country. For fourteen of the core concepts, respondents were asked—using eleven-point unipolar scales—how important they felt the concept was for democracy in general (hereafter "importance" items), followed by the extent to which they felt the concept applied in their own country currently (hereafter "evaluation" items). For two of the core concepts, *opportunities for effective immigrant participation* and *horizontal accountability*, evaluation items were not included in the final module due to high item non-response found in pre-testing (both the omnibus surveys and in the pilot survey; see Appendix A for details). For the remaining three concepts, *type of governmental coalition*, *responsiveness to the citizens*, and *freedom of expression*, pre-testing results indicated a clear conceptual dichotomy, whereby respondents should first be asked to express their preference (e.g., for single-party governments vs. coalition governments) before being asked importance and evaluation questions that were tailored to their initial preference (hereafter "trade-off" items).

FOCUSING ON IMPORTANCE FOR DEMOCRACY

In the initial stages of development of the module, different formulations to measure the "importance" items were considered. There were concerns that asking respondents up to twenty-five questions in the same format ("how important is x for democracy"), followed by another twenty-five questions in an identical format ("to what extent does x apply in country"), might lead to respondent fatigue, straight-lining, satisficing, or worse, interview break-offs. However, in order to allow for the calculation of scores for satisfaction or dissatisfaction with democracy in a respondent's country according to what they believed to be important, it was necessary for "importance" to be consistently measured for all aspects of democracy.

The formulations that were considered for the importance items were sourced from previous surveys as well as suggestions by the Question Module Design Team (QDT) and Core Scientific Team (CST). The formulation "how important" was proposed by the QDT and had previously appeared in the ESS (2002), PEW (2009),[1] International Social Survey Programme (ISSP) (2004), "PARTIREP" (2009),[2] EU-Profiler (2009),[3] World Values Survey (2005), and the Comparative National Elections project (CNEP) (2004).[4] "How essential" was another formulation proposed by the QDT (and had previously appeared in CNEP), and "how necessary" and "how acceptable" were proposed by the CST as potential alternatives.

"Important" was the adjective chosen over "essential," "necessary," and "acceptable" because it was felt to be a closer match to what the QDT were intending to measure. The

[1] PEW Research Center, United States, <http://www.pewresearch.org/>.

[2] <http://www.partirep.eu/>.

[3] <http://www.eui.eu/Projects/EUDO/Research/EUProfiler.aspx>.

[4] <http://www.cnep.ics.ul.pt/index1.asp>.

focus is on the value placed on features of a democracy by the respondent. In British English, the terms "essential," "necessary," and "acceptable" can be used instead of "important" and the question would still make sense. However, by using these terms (rather than "important") we would not have been measuring the same thing. The word "essential" is stronger than "important" but does not capture the notion of "value" in the same way that "important" does. The word "necessary" implies that something is required, but again, the notion of how significant this is (or its value) is missing; "acceptable" implies something is "good enough" or "satisfactory" but nothing better than that and again, the idea of value is missing.

In order to facilitate analysis of the module, the QDT were keen to ensure that the same structure was chosen for both the importance and the evaluation items as far as reasonably possible. In other words, if an eleven-point scale was used to measure responses to an importance item, then an eleven-point scale should also be used to measure responses to an evaluation item. Similarly, the CST wanted to ensure that the scale labels chosen were similar enough conceptually to combine within the module (to avoid combining "importance" with "acceptability," for example). This would also avoid potential problems of equivalence across countries once the questions had been translated. At the same time, the CST also wanted to take measures to alleviate respondent burden and potential problems of satisficing (Krosnick 1999). In the end, almost all of the importance questions were measured using the formulation "How important is X for a democracy in general," and the responses were measured using an eleven-point scale labeled as "Not at all important for democracy in general (0) – Extremely important for democracy in general (10)."

An additional concern related to "importance" was to convey to respondents that the questions focused on importance for *democracy* rather than a *general* sense of importance for society. Feedback from the pilot interviewers in Great Britain and Russia, and from respondents debriefed in each country, revealed that respondents did not always focus on

what is important *for democracy*, but instead thought about what they personally considered to be important in life generally. To address this, the response scale labels for all of the importance items in the module were amended to include the phrase "for democracy in general." This appeared on the showcards, helping to reinforce the link with democracy for respondents.

In addition, feedback from respondents who were debriefed as part of the pilot survey indicated that they found it difficult to answer evaluation items about "governments in [country] in general" and some respondents thought about the current government when answering. There were concerns that in the mainstage survey this might have been a particular issue in countries where respondents felt very differently about the current government compared to past governments. To pre-empt the inconsistencies this may have generated, respondents in the mainstage survey were asked to think about how "democracy is working in [country] today."

DECIDING WHETHER TO ADMINISTER IMPORTANCE AND EVALUATION ITEMS "PAIRWISE" OR "LISTWISE"

The second challenge of designing a module on democracy concerned the order of the questions and whether to present the importance and evaluation items in pairs (according to the concepts measured) or in two separate lists (with all importance items asked first and all evaluation items asked afterwards).

To assess the impact of question order on response, a selection of items were tested on face-to-face omnibus surveys in Hungary (N = 1,046), Portugal (N = 1,263), and the UK (N =

1,002) in May–June 2011. This pre-test included a split-ballot experiment, whereby respondents were randomly allocated to one of two groups. In one group, the importance and evaluation items were administered in pairs ("pairwise"), whereby each importance item was directly followed by its corresponding evaluation item. In the second group, the items were administered "listwise," that is, ten importance items were administered in a battery formulation, followed by ten evaluation items in a separate battery. For each concept, an eleven-point scale from zero to ten was used, where the end point labels were tailored to each importance and evaluation question.

The experiment aimed to explore whether the two ways of arranging the importance and evaluation items had differing impact on indicators of satisficing. Satisficing can occur when survey respondents are not motivated to carefully consider a question before responding, when the task of responding is too difficult because of the language used or the cognitive effort required, or when they tire of answering questions that use the same response scale or similar formats. Any one of these factors may lead respondents to engage in shortcuts when answering. Indicators of satisficing may include frequent use of scale mid-points or extreme end points, non-differentiation between the answers given to different items, straight-lining (whereby respondents give the same answer to all items asked in a set), or tendencies to give "non-answers" such as "don't know" or refuse to answer at all (Kaminska, McCutcheon, and Billiet 2010; Krosnick 1991, 1999).

The data from the omnibus surveys revealed evidence of frequent use of scale mid-points and extreme end points, non-differentiation between the importance and evaluation items, and high item non-response, but not of straight-lining. *Table 2.1* shows the percentage of respondents who scored at either of the two extreme points of the scale (0 and 10), who scored at the mid-point (5), or who answered "don't know," for each experimental condition and for each of the twenty items.

Use of extreme points of the scale by a respondent might be considered a "weak" form of satisficing. Respondents must think in sufficient detail about a question to determine the end of the scale at which their opinion lies, but choosing the extreme point enables them to avoid the additional cognitive effort required to differentiate between adjacent points on the scale. Use of the mid-point is also an indicator of satisficing as a means of avoiding "taking sides" (Krosnick and Fabrigar 1997). The use of "don't know" is considered an indicator of "strong satisficing" if a respondent actually avoids any judgment at all (Krosnick 1991).

<COMP: INSERT TABLE 2.1 NEAR HERE>

It is possible to test whether the responses to the questions asked in the omnibus surveys might be the result of respondents becoming fatigued or disinterested when responding to the battery of items. If this were the case, questions administered later in the battery would exhibit more evidence of satisficing than questions administered earlier. In the listwise condition, increased satisficing would be expected to occur in the evaluation items, as they were administered later in the battery. In the pairwise condition, increased satisficing would be expected in the responses to both the importance and the evaluation items making up the pairs asked later in the questionnaire. These patterns can be seen in the data to some extent in the use of "don't know" (see *Table 2.1*).

Generally speaking, in all three countries, "don't know" was used more when the items were administered listwise rather than pairwise. It is possible that alternating the importance and evaluation items through pairwise administration prevented fatigue. A regression analysis (controlling for country and item placement within the module) demonstrated that "don't know" answers were significantly less likely when questions were asked pairwise than listwise, and significantly more likely the later the questions were asked in the module (Martin 2011) (see Table 2.2).

Use of the extreme scale points (0 and 10) were not significantly influenced by experimental condition or position in the questionnaire. However, items administered pairwise were significantly more likely to elicit use of the mid-point than those administered listwise (Table 2.2).

<COMP: INSERT TABLE 2.2 NEAR HERE>

In all three countries there was less differentiation, that is, a higher percentage of respondents giving the same scores, between importance and evaluation items when the items were administered pairwise compared to when they were administered listwise (*Figure 2.1*). For all questions tested in the experiment, a significantly higher percentage of respondents gave the same answer to the importance and evaluation items (within a concept) when they were asked in pairs (with the evaluation item asked immediately after its corresponding importance item) compared to when they were asked listwise (see *Figure 2.1*).

<COMP: INSERT FIGURE 2.1 NEAR HERE>

The increased non-differentiation when items were administered pairwise suggested that respondents were failing to distinguish between the *importance* and *evaluation* questions. The distinction between these two types of question was a critical feature of the module. Therefore, despite the increased risk of item non-response generated by the listwise condition, it was decided that the importance and evaluation items should be administered listwise due to the greater levels of differentiation demonstrated. This is one example of the trade-off between different forms of measurement error (item non-response and non-differentiation) the CST and QDT had to make during the design of the module.

ABSTRACT CONCEPTS AND INTERPRETATIONS OF "DEMOCRACY"

A module of questions on democracy necessarily covers some topics seldom considered by many besides the most politically engaged members of the general population. Thus, it is not possible simply to ask respondents for their opinions on an abstract concept such as *horizontal accountability* directly, for example, "How important is horizontal accountability in a democracy?"; instead, questions must be formulated in a way that respondents can comprehend (Zaller and Feldman 1992). In other words, a balance between theoretical concepts and everyday terms and ideas must be achieved in order for the questions to be answerable by all respondents in all countries and to avoid respondents thinking that their knowledge is being tested.

Cross-national equivalence can be threatened by varying interpretations of a question in the different countries in which it is fielded. To understand how key terms in ESS questions are interpreted cross-nationally, cognitive interviewing is included as a qualitative stage of pre-testing. Cognitive interviewing in the ESS involves asking respondents a test question, then asking a series of standardized probes to explore their understanding of key terms, how easy or difficult they found the question to answer, how they reached their answer, and how they interpreted different points on the answer scale (see for example Miller et al. 2005; Miller et al. 2011). Issues—or errors—with a question identified through cognitive interviewing can be classified according to the Cross National Error Source Typology (CNEST) (Fitzgerald et al. 2011). Errors may be due to poor source question design, translation problems resulting from translator error, translation problems resulting from features of the source question that make translation difficult, or problems with cultural

portability, whereby either a concept does not exist in all countries, or it exists in a way that prevents the proposed measurement approach from being used.

A selection of eight items from the module were tested using cognitive interviewing, to try to establish how respondents interpreted and understood the questions. Ten interviews per country were conducted in Austria, Bulgaria, Israel, Portugal, and the UK, with respondents selected according to quotas based on age, gender, education level, and level of interest in politics.

After respondents were asked the first democracy question, they were probed on what "democracy" meant to them when answering the question. There were respondents in all countries that associated democracy with elections or "people power." In some countries there were also references to "not being a dictatorship," "freedom," "freedom of speech," and "equal treatment." There were also respondents in all countries who found it impossible to articulate what "democracy" meant to them when answering. This may reflect that democracy is such a widely accepted concept that it is not possible for some people to articulate what it is in overarching terms, but they are still able to answer more detailed questions about democracy.

The analysis of data from the cognitive interviewing revealed several issues with the items tested. Problems with the source question design included differing interpretations of "equal treatment by the law" and whether this referred to the courts or to how the law was written (*Accessibility and equality of the judicial system*).[5] This was resolved by making the question more specific in referring directly to "the courts." Another issue was that the phrase "deciding major issues by voting directly in referendums" was found to be confusing for

[5] Cognitive interview question wording: "How important would you say it is for a democracy that everyone is treated equally by the law? Choose your answer from this card where 0 is not at all important and 10 is extremely important."

some respondents (*Forms of participation*). To overcome this, "deciding major issues" was replaced with "having the final say on the most important political issues." The analysis of the cognitive interviewing data also revealed issues with the response scale when each end point represented opposing positions. Respondents were sometimes confused as to how the mid-point of the scale should be interpreted (*Type of electoral system*).[6] This was resolved by the introduction of "forced choice" questions with tailored follow-up items, which is discussed later in this chapter.

Some problems with translator error were also detected, although no issues were classified as "translation problems resulting from poor source question design." The Bulgarian translation of "national elections" excluded the term "national," leading respondents to think about all elections, including those for private members clubs (*Free and fair elections*). The reason for focusing on "national elections" was emphasized to all National Coordinators (NCs); in addition, an annotation for "national elections" was added to the final questionnaire: "'national elections' refers to national elections for a country's primary legislative assembly."

Finally, some issues relating to cultural portability were identified. There was wide variation in the types of "minorities" respondents were thinking about across different countries (*Subjects of representation*). This, combined with poor performance in other stages

[6] Cognitive interview question wording: "Some countries have a system for national elections that generally results in one party winning and forming a government on its own. Other countries have an election system that generally results in more than one party forming a government and sharing power. I now want to ask which system you think is better for a democracy regardless of the system used in your own country at present. Use this card where 0 means a system which generally results in one party forming a government and 10 means a system which generally results in more than one party forming a government."

of pre-testing, led to the concept being dropped from the module. There were also respondents in Bulgaria and Austria who expressed a lack of understanding of how the courts might "overrule governments that abuse their powers" (*Horizontal accountability*). In Bulgaria this was thought to be related to the rarity of this happening in the country, whereas in Austria the lack of specific reference to the "constitutional court" was felt to be problematic. Although the question wording was somewhat improved as a result,[7] a compromise had to be made in retaining the general focus of the question on "the courts" rather than permitting a reference to the "constitutional court" to be included only in countries where these exist.

CULTURAL EQUIVALENCE AND TRANSLATION

Designing a questionnaire that is functionally equivalent is a key element of enabling comparisons to be made cross-nationally (Johnson 1998). Certain issues relating to democracy may be considered less relevant to some countries, for example, a question about the importance of referenda or coalition governments may seem meaningless to respondents in countries where there are rarely referenda (e.g., Israel, Russia, or the UK) or almost always single-party governments (e.g., Spain). To address this, a question that is general enough to cover all country-specific options but accompanied by an unfortunate loss in detail could be asked, or the item could be adapted into country-specific questions that cannot be directly compared (Smith 2004). For the ESS democracy module, respondents in all countries were

[7] From (Cognitive interview question wording): "How important would you say it is for a democracy that the courts are able to overrule governments that abuse their powers?" To (Final question wording): How important would you say it is for a democracy in general that the courts are able to stop the government acting beyond its authority?

asked the same questions,[8] even when relevancy was considered to be low. The issue of low relevancy was considered unlikely to apply to many items or many countries, and this approach, combined with making items slightly more general where necessary, was a sensible compromise.

For a question to be comparable in different countries, direct word-for-word translations are not always necessary, or indeed possible (Harkness 2007; Harkness, Edwards, Hansen, Miller, and Villar 2010). The ESS never insists on word-for-word translations but does require the same direct stimulus to be provided to all respondents. Where the meaning of a word used in the source questionnaire (in British English) is unclear, annotations are provided in the form of footnotes. These are not intended to be "incorporated literally into translated questions, nor provided to interviewers as notes" (Harkness 2007: 88), but ". . . simply to be used as aids to the design of functionally equivalent [translated] questions" (Harkness 2007: 88). In the democracy module, ambiguous terms that may have been interpreted differently in other languages were annotated. For example, the phrase ". . . are free to criticise" was annotated as "are free to" in the sense of "are allowed to" and "criticise" in the sense of "contest or dispute" rather than "being able to disrupt."

ENSURING VARIATION IN RESPONSE TO CONSENSUAL ITEMS

[8] With the exception of the item measuring viable opposition—". . . different political parties offer clear alternatives to one another"—where countries were permitted to change "political parties" to "candidates" instead of or in addition to "political parties" if appropriate.

A further area of consideration when designing the module was the need to ensure variation in the responses given to consensual items. These are questions which focus on particularly salient or fundamental democratic concepts. It has been argued that democracy can be considered a universal value (Sen 1999). As such, one might expect skewed responses to consensual items such as *free and fair elections, accessibility and equality of the judicial system*, or *freedom of speech*. The use of eleven-point scales to measure perceived importance of these concepts can go some way to increasing variation in responses (Krosnick and Fabrigar 1997).

However, during pre-testing and piloting, responses to the questions about the importance of free and fair national elections and equal treatment by the courts were skewed towards the upper end of the scale. Whilst this was felt to be unavoidable to a large extent in these concepts, it was possible to reduce it in others by introducing the idea of extremes. Taking freedom of speech as an example, variation in responses was increased by asking respondents about freedom of expression of political views "even if they are extreme." An earlier version of the question wording referred to opinions that were "damaging for the government," which was felt to be too vague. Sniderman et al. (1996) found that only a minority of 35 percent of respondents support freedom of speech for specific groups that they particularly dislike. The format for the most disliked groups in this study was taken from Sullivan, Piereson, and Marcus (1979: 793), who asked the following question:

> I am giving you a list of groups in politics. As I read the list please follow along: socialists, fascists, communists, Ku Klux Klan, John Birch society, Black Panthers, Symbionese Liberation Army, atheists, pro-abortionists, and anti-abortionists. Which of these groups do you like the least? If there is some group that you like even less than the groups listed here, please tell me the name of that group.

The QDT suggested that the list of least liked groups to use in the ESS could be: Fascists, Communists, Islamists, atheists, feminists, racists, immigrants, and drug addicts. The CST felt that a generic list of least-liked groups might not be equivalent across all countries and the challenges in formulating a list that would be inclusive enough to be relevant cross-nationally were felt to be insurmountable. The final question simply asked about freedom of speech for those with extreme political views, thereby making it easier for respondents to give a larger range of scores, increasing the variation.

SOCIAL DESIRABILITY

Another difficulty in measuring understanding and evaluations of democracy in a cross-national survey relates to social desirability bias. Social desirability is ". . . the tendency of individuals to 'manage' social interactions by projecting favorable images of themselves, thereby maximizing conformity to others and minimizing the danger of receiving negative evaluations from them" (Johnson and van de Vijver 2002: 194). A respondent may give a socially desirable response in order to present themselves as "being better or more capable than others" or to try "to harmoniously fit in and gain social approval" (Lalwani, Shavitt, and Johnson 2006: 166).

Social desirability can create response bias and pose a serious threat to the validity of research findings (DeMaio 1984). However, social desirability does not affect the accuracy of *all* responses to *all* survey questions equally. Johnson and van de Vijver (2002) and Streb et al. (2008) have indicated that some questions are more susceptible to socially desirable answers than others, for example, those asking about socially sensitive or controversial issues. Some modes of data collection are also more likely to generate socially desirable responses than others, for example, those that are less anonymous, such as face-to-face

interviewing (Johnson and van de Vijver 2002). Similarly, some respondents are more likely to exhibit socially desirable behavior than others depending on their cultural background (Lalwani, Shavitt, and Johnson 2006).

Within the democracy module, the potential for respondents to give a socially desirable response was high. Researchers have previously struggled to overcome this issue when measuring support for democracy. Moreno and Méndez (2002) argue that "a good assessment of democratic values should not be limited to measuring overt support for democracy, because democracy has become a concept affected by social desirability bias" (2002: 365).

Unfortunately, "there is no simple safeguard against social desirability" (Johnson and van de Vijver 2002: 202). However, by introducing some basic measures into the module, the potential for eliciting socially desirable responses was reduced. These measures included informing respondents that there are no right or wrong answers. This approach was used at the beginning of the module (preceding questions about the importance of different aspects of democracy), halfway through (preceding questions measuring evaluations of democracy), and towards the end of the module (before the forced choice questions). The CST hoped this approach would encourage honest responses but it was not possible to check the independent effect of this. A second measure to address potential social desirability bias was to emphasize the general nature of the questions. There was also a focus on general concepts to avoid specific issues that could have been considered politically sensitive in one or more participating countries, as well as a focus on concepts that were applicable across all participating countries (regardless of whether the democracy was "old" or "new"). The words "in general" were included on scale labels to emphasize that this was the focus of attention (rather than the respondent's own country). It was also hoped that asking respondents about different elements that could be considered as important for a democracy would produce a

more nuanced picture of attitudes rather than only measuring direct, overt support of democracy.

Care was taken when wording the questions to avoid leading respondents towards a particular response. Furthermore, the use of eleven-point scales in the majority of the questions, aimed to encourage greater differentiation in responses, and the inclusion of a small number of "forced choice" questions facilitated the expression of opposing views.

It might be expected that respondents in more authoritarian (and/or newer) democracies would be more likely to give socially desirable responses compared to those in established democracies. Indeed, Inglehart and Welzel argued that "empirically, the Albanians and Azerbaijanis are more likely to say favourable things about democracy than are the Swedes or the Swiss" (Inglehart and Welzel 2004: 9). This was felt to be explained partly by the emergence of "Critical Citizens" among those in stable democracies,[9] and partly because "at this point in history, saying favourable things about democracy has become the socially desired response in most societies" (Inglehart and Welzel 2004: 9). Due to the need to retain cross-national equivalence in all participating countries (regardless of the democracies within each), changes that could have been made to overcome this concern (such as amending the question wording or adding instructions for specific countries) were not made. Instead, this was taken into consideration during fieldwork. For example, in Albania interviewers highlighted that the survey was not commissioned by the government in order to reassure respondents of the independent nature of the survey, thus encouraging participation and trust towards the interviewers.

[9] A term coined by Norris (ed.), 1999.

FORCED CHOICE QUESTIONS (DICHOTOMOUS "TRADE-OFFS")

This final section summarizes one of the most challenging parts of designing the democracy module—how to measure differing points of view about an issue.

Unipolar scales were appropriate to measure many of the concepts in the module (i.e., "how important would you say it is for democracy in general that . . ."). This format assumes, to some degree, that the respondent has an overall basic acceptance of the concept being asked about. For example, the likelihood of respondents holding the view that it is important for democracy *not* to have free and fair elections would be low. Therefore, it would be appropriate to measure this concept with a question that assumes acceptance of the concept, that is, "how important would you say it is for democracy in general that national elections are free and fair?" In later discussions between the QDT and CST during the design process, concerns were expressed that for a number of the importance items, it would not be sufficient to provide respondents with the option of evaluating the concept as "not important for democracy" when a substantial proportion of them may hold the opinion that it is the opposite of the statement that is extremely important for democracy. In these cases, a unipolar scale would be inappropriate.

Take, for example, an earlier version of the question measuring *responsiveness to the citizens*:

When there is disagreement, should the government in a democracy rather follow the views of the citizens, or should they rather follow their own judgment? Please use this card.

Government should still follow public opinion									**Government should use their own judgement**		**(Don't know)**
00	01	02	03	04	05	06	07	08	09	10	88

Following poor performance of similar bipolar items in pre-testing—whereby confusion around the meaning of the mid-point seemingly led to high item non-response and poor quality scores in Survey Quality Predictor (SQP) coding—attempts were made to formulate a unipolar item:

How important would you say it is for a democracy that governments follow public opinion, *even* when they disagree with it?

Not at all important									**Extremely important**		**(Don't know)**
00	01	02	03	04	05	06	07	08	09	10	88

When asked as a unipolar question, this item may be problematic for some respondents who feel that it is actually extremely important for a democracy that governments do *not* follow public opinion if they disagree with it. In this case, choosing an answer from the "not at all important" end of the scale does not accurately reflect one's opinion. When survey questions are poorly designed in such a way that respondents are unable to choose an answer that reflects their view, or are required to expend extensive cognitive effort, they might be more

inclined to give a "don't know" response or satisfice in some other way, such as choosing the mid-point of the scale or selecting a response at random (Krosnick 1999).

In an attempt to address this issue, the CST proposed a series of dichotomous questions, whereby respondents were first asked to choose between two clear alternatives in regard to a concept (with "it depends" as a hidden code), followed immediately by a question that asked them to assess how important their selected choice was for a democracy. This would be less cognitively demanding for respondents and would present them with an importance item relevant to their viewpoint.

Seven dichotomous questions were included in the main pilot questionnaire for six concepts: *Forms of participation* (referendum), *Type of electoral system* (proportional representation/majority representation), *Responsiveness to citizens* (public opinion), *Responsiveness to other stakeholders* (business), *Freedom of expression* (those holding extreme political views), and *Subjects of representation* (majority/take account of minority groups and immigrants voting).

Each of these dichotomous ("forced choice") items was then followed by a question asking about the importance of the selected choice for a democracy, for example:

C10 CARD 32 Countries differ in whether their governments are generally formed by a single party or by two or more parties. Which **one** of the statements on this card describes what you think is generally better for a democracy? Would you say that . . . **READ OUT . . .**

INTERVIEWER: CODE ONE ANSWER ONLY

In a democracy governments should generally be formed by:

A single party	1	**ASK C11**
Two or more parties	2	
(Neither of these/it depends)	5	**GO TO C12**
(Don't know)	8	

ASK IF CODE 1 or 2 AT C10

C11 CARD 33 And how important do you think it is for a democracy that governments are generally formed by [a single party/two or more parties]?

INTERVIEWER NOTE: Read text in brackets according to answer given at C10.

Not at all important										**Extremely important**	**(Don't Know)**
00	01	02	03	04	05	06	07	08	09	10	88

In addition, all respondents were also presented with an alternative version (to test which worked better) whereby they were asked a single question about the importance of only one "side" within each dichotomous question. This single item importance (with a single item evaluation) format was used in the pilot for all other concepts.

The dichotomous format required an additional question item for each relevant concept. There were also concerns that a large number of respondents would choose the "it depends" category, which left no viable follow-up question.[10]

Data from the pilot study in the UK and Russia revealed that item non-response was problematic, particularly in Russia (see *Table 2.3*). Despite the high item non-response in some cases, the data also clearly show that—for many of the concepts—there were respondents who held views on each side of the dichotomy. The dichotomous format was therefore retained for three concepts: *type of governmental coalition*, *responsiveness to the citizens*, and *freedom of expression*. However, two concepts—*responsiveness to other stakeholders (business)* and *subjects of representation (majority/take account of minority groups)*—were considered lower priority for the module and were ultimately excluded from the final module due to extremely high item non-response.

<COMP: INSERT TABLE 2.3 NEAR HERE>

The concept *subjects of representation (immigrants' right to vote)* was measured in the final module with an importance item only due to very high item non-response for the evaluation item. For the three concepts measured in the final module with a dichotomous preference item and tailored importance items, the question wording was simplified, clearer introductions were added to the questionnaire to signpost respondents to the changes in question format, and to the changes between importance and evaluation items in order to limit the possibility of non-differentiation in responses. Tailored evaluation items were also introduced for these three concepts, with respondents who gave a "don't know" or "it depends on the circumstances" response to the dichotomous preference item routed to the

[10] A question on the importance for a democracy that "it depends whether x or x happens" would be too complex in a standardized interview.

tailored importance and evaluation items most prevalent in the UK pilot in order to limit item non-response.

The need to introduce questions with dichotomous "trade-offs" was one of the most difficult parts of the democracy module to design. For the concepts affected, two clear points of view about an issue existed and it was not appropriate to make respondents answer a single question about only one of these perspectives. The solution enabled respondents to answer a reasonable question and convey their viewpoint—thereby meeting the concerns held by the CST. At the same time, the question and response format complemented the other measures in the module and provided responses that could still be used in analyses—thereby meeting the needs of the QDT and other potential data users.

CONCLUSION

The ESS employs one of the most thorough question design processes of any social survey and this is a key feature of its rigorous methodology. Collaborators with different interests work together to design each question module: the Question Module Design Team, whose priorities lie in testing their academic theories; the ESS National Coordinators, whose main role is to advise on how a question might perform in their language and cultural context; and the ESS Core Scientific Team, who must balance the need to focus on how a question can be developed to be understood equivalently across multiple languages and cultural contexts with the desires of the Question Design Teams.

Designing a module of questions for the European Social Survey involves several types of compromise. A balance must be achieved between attempts to perfectly capture the theoretical construct discussed in an academic field (such as *rule of law: accessibility and equality of the judicial system*) and how the concept can practically be measured in a survey.

This is often achieved by simplifying the language used or only focusing on one aspect of a concept in order to make it meaningful to the general public (e.g., asking only about "the courts" rather than "the law," which is more open to interpretation). Sometimes, a question must be altered in order to produce useful data, for example, asking about freedom of speech for those with extreme political opinions in order to avoid producing a variable in which almost all respondents give the same answer (*freedom of speech*). Other compromises must be made where a specific question might work very well in some countries, such as asking about the constitutional court in countries where there is one, but needs to be made more general (and therefore less straightforward in those countries) in order to be fielded in *all* participating countries, for example, asking about "the courts' ability to stop the government acting beyond its authority" (*horizontal accountability*).

Compromise is also necessary when considering the different types of measurement error than can have an effect on specific questions. For example, the decision to administer the importance and evaluation questions as a list (rather than as pairs) was largely based on the lower levels of non-differentiation demonstrated by this approach during pre-testing. The fact that this approach also generated greater item non-response was deemed acceptable in order to create a module that met the measurement aims of the QDT, NCs, and CST.

Despite the challenges involved, with the necessary compromises made, a forty-one-item module on democracy was successfully implemented in twenty-nine countries in ESS Round 6. The average item non-response rate for all countries, across the whole module, was just 4.2 percent.[11]

[11] Excluding the items: "How important is it for you to live in a country that is governed democratically? Choose your answer from this card …." and "How democratic do you think [country] is overall? Choose your answer from this card …."—for which the item non-response was below 4%.

The data and associated documentation are freely available from <http://www.europeansocialsurvey.org/>.

Figure 2.1 Percentage of equal score[12] across questionnaire versions (% valid cases)

N = 2,591–3,209 (differs across concepts), differences for each item are significant at $p < .01$.

Table 2.1 Percentage of respondents who chose extreme scores, the mid-point, or "don't know." UK, Hungary, and Portugal

		Extreme scores (0 or 10), %		**Mid-point (5), %**		**Don't know, %**	
		Pair	List	Pair	List	Pair	List
Importance items	Accessibility and equality of the judicial system	59.7	65.0	4.9	3.4	1.6	1.4
	Forms of participation	31.0	38.9	11.6	8.2	3.6	4.4
	Freedom of press	39.0	46.7	10.9	7.5	2.5	4.0
	Viable opposition	38.6	45.1	7.7	6.5	2.7	4.2
	Horizontal accountability	38.0	44.3	11.3	7.8	5.7	5.5
	A particular minority in society	33.9	41.9	11.4	9.1	4.0	3.6
	Opportunities for effective participation	22.7	31.5	14.0	12.6	7.3	6.6
	Type of electoral system	27.8	25.2	12.9	12.4	8.8	9.3
	Subjects of representation	24.7	22.4	15.8	12.4	6.0	5.4
	Efficiency	21.0	19.5	14.5	14.7	8.8	9.7

[12] "Equal score" refers to the same score for the meaning item and corresponding evaluation item.

Evaluation items	Accessibility and equality of the judicial system	20.7	19.7	17.0	14.4	2.5	3.9
	Forms of participation	16.7	15.4	18.0	15.5	6.2	9.4
	Freedom of press	18.5	19.4	16.7	13.2	4.0	6.8
	Viable opposition	27.9	25.8	12.2	13.1	3.9	5.8
	Horizontal accountability	13.3	14.3	17.0	16.0	9.9	14.1
	A particular minority in society	12.8	11.3	14.0	18.1	5.9	7.9
	Opportunities for effective participation	9.3	10.0	16.2	14.8	23.0	25.7
	Type of electoral system	13.2	12.4	22.9	21.6	11.7	14.4
	Subjects of representation	11.8	9.4	18.7	18.3	9.1	10.2
	Efficiency	10.6	10.2	18.5	16.2	18.2	22.2

N = 1,661

Table 2.2 Unstandardized regression coefficient predicting the percentage of "don't know" answers and the scores 0, 5, and 10 for 20 items as a function of experimental condition (listwise or pairwise) and position in the questionnaire (1–20)

Response	Predictor	Coefficient
Don't Know	pairwise	-1.501**
	position	.183**
Score 0	pairwise	.001
	position	.101
Score 5	pairwise	1.271*
	position	.044
Score 10	pairwise	-1.565
	position	-.332

*$p < .05$; **$p < .01$; since each of the twenty items is asked in two conditions and in three different countries, N = 120 items (20*2*3). Control variables: Country (dummies for two countries), and all items (19 item dummies).

Table 2.3 Frequencies (%) of dichotomous items in the ESS6 pilot, N = 823

Concept	**Option 1**	**Option 2**	**Country**	**Option 1 (%)**	**Option 2 (%)**	**Neither of these/it depends (%)**	**Don't know/refused (%)**
Forms of participation (referendum)	Parliament	People—by voting on them directly in referendums	UK	29.2	66.3	3.0	1.5
			RU	17.5	67.3	8.3	6.9
Type of governmental coalition	A single party	Two or more parties	UK	48.6	45.9	3.5	2.0
			RU	24.2	60.4	8.3	7.1
Responsiveness to the citizens	Change their policies and plans in response to public opinion	Stick to their policies and plans regardless of public opinion	UK	73.1	17.7	7.5	1.7
			RU	70.4	11.4	11.8	6.4

Responsiveness to other stakeholders (business)	Change their policies and plans in response to business demands	Stick to their policies and plans regardless of business demands	UK	48.9	35.2	10.0	6.0
			RU	43.4	21.1	20.4	15.2
Freedom of expression	Everyone **should** be free to express their political views openly, even if they are extreme	Those who hold extreme political views should **not** be free to express them openly	UK	73.6	19.2	5.5	1.7
			RU	43.1	36.0	11.4	9.5
Subjects of representation (majority/take account of minority groups)	Governments should only follow the demands of the majority	Governments should take into account the demands of minority groups as well	UK	18.5	76.1	3.0	2.5
			RU	26.1	57.1	12.1	4.7

Subjects of representation (immigrants' right to vote)	Immigrants should get the right to vote in national elections even if they are not citizens of that country	Immigrants should get the right to vote in national elections only when they become citizens of that country	UK	6.5	88.0	5.0	.5
			RU	11.6	65.4	13.5	9.5