ISER Working Paper Series

Data Quality in Telephone Surveys and the Effect of Questionnaire Length: a Cross-National Experiment

Caroline Roberts
Swiss Centre of Expertise in the Social Sciences (FORS)
Université de Lausanne

Gillian Eva
Marie Stopes International
London, UK

Nick Allum
Department of Sociology
University of Essex

Peter Lynn
Institute for Social and Economic Research
University of Essex

No. 2010-36
9 November 2010

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Non-Technical Summary

In survey interviews carried out by telephone, it is possible that respondents may get tired, bored, or generally less attentive as the interview goes on. If this were the case, then answers given to questions later in the interview should be of lower quality than answers given to questions earlier in the interview. This study assesses whether we do indeed find lower quality answers when questions are asked later in an interview. The study used data from an experiment in which questionnaires of three different lengths were used. The experiment took place in Germany, Hungary, Poland and Switzerland, with interviews lasting approximately 30, 45 or 60 minutes. The questionnaires were based on the European Social Survey.

The main conclusion is that we do indeed find evidence that data quality is lower when questions come later in the interview. The final section of the paper discusses some practical implications of this finding for survey designers.

Data Quality in telephone surveys and the effect of questionnaire length: a cross-national experiment

Caroline Roberts
Swiss Centre of Expertise in the Social Sciences (FORS), Université de Lausanne

Gillian Eva
Marie Stopes International, London UK

Nick Allum
Department of Sociology, University of Essex

Peter Lynn
Institute for Social and Economic Research, University of Essex

Abstract: Respondents in long telephone survey interviews may adopt satisficing strategies as they approach the end of the questionnaire (Holbrook, Green and Krosnick, 2003). However, there is inconsistency regarding the relationship between questionnaire length and different forms of satisficing. We investigate whether long questionnaires are associated with a reduction in response quality using data from a cross-national survey experiment. Sample members were randomly assigned to interviews of 60, 45 or 30 minutes. We compare responses to attitudinal measures from a module on happiness and well-being, which was asked at different points in the interview in each of the three groups.

Keywords: measurement error, respondent fatigue, prior questions, satisficing, survey error

JEL Codes: C81, C83

Acknowledgments: The study reported here formed part of a programme of methodological research into data collection modes managed by the European Social Survey (ESS) Central Co-ordinating Team at the Centre for Comparative Social Surveys (CCSS), City University London and funded by a European Commission FP7 infrastructure grant for "Integrating Activities" (2006-2010). The first and second authors were researchers at CCSS at the time the study was designed and implemented. The fourth author is adviser to the ESS programme of research into data collection modes and he is grateful for the support of the ESRC UK Longitudinal Studies Centre. Fieldwork in Hungary was co-financed by The Gallup Organisation, Europe and fieldwork in Switzerland was co-financed by the Swiss National Science Foundation. The authors wish to thank the ESS National Coordinators and their colleagues who were responsibile for coordinating data collection and delivery and liaising with fieldwork agencies in each of the participating countries: Jan van Deth at the University of Mannheim in Germany; Agnes Illyes at The Gallup Organisation in Hungary, Pawel and Franek Sztabinski and Teresa Zmijewska-Jedrzejczyk at the Centre for Social Survey Research, Institute of Philosophy and Sociology, Polish Academy of Sciences in Poland; and Dominque Joye and Nicole Schoebi at SIDOS (now FORS, the Swiss Centre of Expertise in the Social Sciences) in Switzerland.

Contact: caroline.roberts@unil.ch

INTRODUCTION

Surveys contain errors of many different kinds, some of which can be easily quantified to provide an indication of the overall quality of the data collected, and some of which are more problematic to detect, measure and control for. Measurement errors in particular present difficult challenges for survey researchers because they can take a variety of forms, and it is not always easy to predict their occurrence. Nevertheless, there is now an extensive literature documenting the different types of response errors that can affect the overall quality of survey data, and showing some consistent patterns about when and where such errors are likely to occur. Groves (1979) has argued that measurement error in surveys can be attributed either to the 'actors' involved in the survey process (notably, the interviewer and respondent in interviewer-administered surveys) or to the 'questions' asked and the way they are administered to survey respondents. In this paper, we focus on how these two elements – actors and questions – interact to produce errors in the data.

In order to understand the types of measurement error we are interested in here, it is helpful to look at the cognitive processes involved in answering survey questions. Tourangeau, Rips and Rasinski's (2000) model of the survey response process (see also Cannell, Miller and Oksenberg, 1981) proposes four main components of processing (each consisting of several sub-components): (1) comprehending the survey question, (2) searching for and retrieving from memory the information requested, (3) formulating a judgement based on the retrieved information, and (4) mapping that judgement on to the available response options in order to select and report an answer. Problems can arise during any of these processes, leading to errors in the data.

Krosnick (1991) proposes a theory about why respondents' answers may contain errors. Executing each of the above stages of processing carefully represents the 'optimal' approach to survey responding and many conscientious respondents may indeed participate in surveys in this way. However, it is likely that for some respondents, the cognitive effort required to complete each of these processes systematically will outweigh the motivation needed to do so. In such situations, respondents will 'satisfice'. The term satisficing was coined by Herbert Simon in relation to more general decision-making processes and combines the terms 'satisfy' and 'suffice' (Simon 1956). In a survey context, it implies that people will expend only sufficient effort for completing the survey task in a way that is

superficially acceptable but that does not represent their 'best' attempt. This may be conscious or unconscious and is manifested by respondents taking shortcuts to reduce the amount of cognitive work involved in the survey task. These shortcuts may take the form of going through each of the necessary processes, but only doing so superficially (referred to by Krosnick as 'weak satisficing'), or it may take the form of skipping processes altogether (referred to as 'strong satisficing').

Different types of errors may be observed, depending on the nature of the shortcutting. For example, weak satisficing includes response effects such as acquiescence, a bias towards agreeing with assertions in the question regardless of content, and response order effects which arise when respondents have a tendency to select the response category that is most accessible in memory – either at the start of a list, where the options are presented visually or at the end of a list where the options are presented orally, regardless of the meaning of the choice (Krosnick and Alwin, 1987). By contrast, strong satisficing includes effects such as repeatedly selecting the 'Don't Know' option, and 'non-differentiation', in which items to be rated on the same response scale are rated on the same scale point (see Krosnick, 1991; Krosnick, Narayan and Smith, 1996; Krosnick, 1999). Other response strategies have also been investigated as possible indicators of satisficing, including selecting the middle response category (O'Muircheartaigh, Krosnick and Helic, 2000) and 'extremeness', a preference for selecting answers from the end points of a scale (Holbrook, Cho and Johnson, 2006).

The likelihood of respondents adopting a sub-optimal response strategy depends on their ability to engage in the necessary processing, their motivation to do so, and the difficulty of the survey task itself. A large number of studies provide evidence consistent with this model (e.g. see Krosnick, Judd and Wittenbrink (2005) for a review). Each factor may be further influenced by other variables in the survey setting. For example, the respondent's ability to expend the required effort may be affected not only by individual factors, but also situational ones, such as the presence of distraction. Motivation to respond 'optimally' may be influenced by the nature of the survey topic, whereas task difficulty (i.e. the cognitive burden of completing the questionnaire) will depend not only on topic, but also on factors such as the types of questions asked, the complexity of question wording, and so on.

Satisficing, then, can be regarded as a putative explanation for a range of response effects observed in surveys where conditions are such that respondent ability and motivation are low and task difficulty is high (Krosnick 1991). Two variables that have been found to influence these conditions are germane to the present investigation: the mode of data collection and the length of the survey questionnaire. Satisficing has been found to be more likely in telephone compared with face-to-face surveys, but this type of mode effect appears to be tempered by the questionnaire length (Jäckle, Roberts and Lynn, 2006; Holbrook, Green and Krosnick, 2003). Longer questionnaires are predicted to be more likely to encourage satisficing because the respondent's motivation typically wanes as he or she progresses through the items (Jabine et al., 1984; p.19; Krosnick 1991; p.224). As the respondent tires, ability to concentrate is also likely to decrease and correspondingly, cognitive burden increases, making shortcutting more likely. Consistent with this, researchers have found evidence of more satisficing on items placed towards the end of the questionnaire (see below). Few studies, however, have explicitly attempted to compare data quality across interviews using questionnaires of different lengths to test the hypothesis that longer questionnaires are more susceptible to response effects than short questionnaires - although see Herzog and Bachman (1981), and Galesic and Bosnjak (2009).

THE PRESENT STUDY

In this paper, we report an experiment carried out in the context of a programme of methodological research on the European Social Survey (ESS). The experiment was primarily designed to investigate some of the specific challenges involved in switching from face-to-face to telephone interviewing, but here we capitalise on features of the design that enable us to examine the effect of varying questionnaire length on the propensity for respondents to satisfice. The analysis employs data from a telephone survey experiment carried out in four countries that participated in round 3 of the ESS: Hungary, Germany, Poland and Switzerland. The principal purpose of the study was to examine the effect of varying the length of the questionnaire on response rates. Sample members (selected using strict probability sampling methods in each country) were randomly assigned to one of three treatment groups, which varied according to the length and design of the questionnaire. Interviewers were instructed to tell the selected target respondent during the survey introduction how long the interview was likely to be. The estimated interview lengths were as follows: group A) 60 minutes; group B) 45 minutes; and group C) 30 minutes. At the end

of their 30-minute interview, group C respondents were asked if they would be willing to participate in a second 30-minute interview, either straight away, or in a separate appointment. This design allowed us to take advantage of the varying length of the questionnaires across experimental groups to examine its effect on the tendency to satisfice. To do this, we use data from a module of questions on psychological and social wellbeing, the placement of which varied in each version of the questionnaire, resulting in a different number of preceding items.

Hypotheses

Our general hypothesis is that the likelihood of satisficing increases with questionnaire length as respondents are more likely to shortcut the response process when motivation is low and task difficulty is high. Motivation is likely to decrease over the course of a long questionnaire, while response burden is likely to increase, so we would expect more satisficing the longer the duration of the interview prior to the target questions being asked.

For the purposes of the present study, we focus on five indicators of respondent satisficing, that have all been employed in previous studies: item non-response, non-differentiation, acquiescence, preference for middle response alternatives and response order effects on rating scales (primacy and recency for fully-labelled ordinal categorical variables and extremeness for scales with end-point labels). The following describes each type of response effect and summarises what is known about how each relates to questionnaire length.

Item non-response

Survey respondents may choose not to respond to certain questions for a variety of reasons, either by simply skipping the question (in a self-completion survey), by explicitly refusing to give an answer to an interviewer, or by giving a 'Don't Know' response. Item non-response has frequently been used as an indicator of poor data quality (de Leeuw and van der Zouwen, 1988), and use of the 'don't know' response alternative has been identified as a form of strong satisficing as it allows respondents to give a legitimate answer without engaging in extensive processing (Krosnick, 1991; Krosnick, 2002). Evidence exists from several studies that item nonresponse is more likely on questions positioned later in the questionnaire (see Krosnick, 2002), although the findings on this have not been consistent and appear to vary by data collection mode. For example, Ferber (1966) concluded that

item nonresponse in a mail survey was not influenced by question position, and Galesic and Bosnjak (2009) found only weak effects of question placement in a web experiment. Consistent with satisficing theory, however, don't know responding has been found to occur more often in (long) telephone interviews compared with face-to-face interviews (Holbrook et al., 2003) and is more likely among respondents with low education (Narayan and Krosnick, 1996).

Non-differentiation

Where a number of questionnaire items are to be rated on the same response scale, respondents may sometimes be tempted to use the same scale-point to rate all (or most) of the items presented in the same set (Krosnick, 1991; p.219). Given that the items in the module analysed here were arranged in blocks sharing a common rating scale we might expect to see evidence of such effects, particularly among respondents in the longer questionnaire group. Consistent with this hypothesis, non-differentiation has, in previous studies, been observed more frequently on sets of items placed later in the questionnaire (e.g. Herzog and Bachman, 1981; Kraut, Wolfson and Rothenberg, 1975).

Acquiescence

Questions with dichotomous response categories of the form 'agree/ disagree', 'true/ false', or 'yes/ no' have been shown to be particularly susceptible to acquiescence bias, in which respondents show a tendency to agree with assertions in the question, irrespective of their content. Respondents answering longer Likert-type scales with agree/ disagree response alternatives are similarly inclined to overuse the 'agree' response (see Saris et al., 2009). Numerous studies provide evidence that the bias results from satisficing, but the findings relating to our specific hypothesis about questionnaire length have been less compelling. For example, Clancy and Wachsler (1971) tested the hypothesis that they would see more acquiescent responses on questions appearing later in the questionnaire due to respondent boredom and fatigue but their prediction was not borne out in the data.

Preference for middle alternatives

The tendency to select the neutral or noncommittal response option in a rating scale has been hypothesised to result from survey satisficing, as it is an easy-to-select, 'easy-to defend' answer for respondents taking shortcuts (Krosnick, Judd and Wittenbrink, 2005; p. 37). There is some debate, however, about whether repeatedly selecting the midpoint

results from respondents shortcutting the response process, whether it reflects a respondents' true ambivalent position or whether some other mechanism (such as social desirability) bias may be at work (O'Muircheartaigh, Krosnick and Helic, 2000; Sturgis, Roberts and Smith, 2010). The use of the midpoint by respondents may depend on how it is labelled (Klopfer and Madden, 1980). For example, Narayan and Krosnick's (1996) re-analysis of Schuman and Presser's (1981) experiments comparing answers to questions with 2 or 3 response options found that respondents with less education were more likely to select the middle alternative, but not on items where the middle category was concerned with maintaining the 'status quo' (see p.75). Nevertheless, consistent with our hypothesis, Herzog and Bachman (1981) found greater evidence of non-differentiation in the middle of the scale on sets of items placed towards the end of long questionnaires. Given the lack of conclusiveness about whether or not midpoint use constitutes satisficing, we include it in this study as an opportunity to gather more evidence.

Response order effects on rating scales

Respondent preferences for the first or last answer category have been attributed to satisficing resulting from the burden placed on processing by long lists of response alternatives. Consistent with this, so-called primacy and recency effects are more common among respondents with less education, particularly on items where respondents must provide answers from a long, unordered list (see Krosnick and Alwin, 1987). However the direction of the effect observed is not always easy to predict, especially where items with rating scales are concerned. Mode of administration provides a clue: where a list of categorical response options is presented visually, primacy effects (preference for the first category) are more likely to occur, and are attributed to the effects of confirmation-biased thinking (*ibid.*). But where such a list is presented orally, as in a telephone survey, both primacy and recency effects have been found to occur, attributed to the competing effects of confirmation bias and limits on the respondent's short-term memory (Krosnick, Judd and Wittenbrink, 2005). With rating scale questions, however, primacy effects have been shown to be more common in both visual and oral modes (e.g. Kalton, Collins and Brook, 1978) because respondents who are satisficing tend to select the first option that even loosely corresponds to their attitude. Based on this, we would expect telephone respondents answering attitude measures with rating scales to show a preference for the first scale point label mentioned by the interviewer. While the weight of evidence supports the conclusion that response order effects are more common under conditions that encourage satisficing

(Krosnick, 1991), there appears to be little to suggest that such effects are more likely for items placed towards the end of the questionnaire (e.g. Carp,1974). The present study provides an ideal opportunity to test this hypothesis anew.

In summary, all response effects that we focus on in this study have been shown in previous studies to be consistent with the theory of survey satisficing, in that they tend to be more common and stronger where there is greater task difficulty, lower respondent motivation and among respondents with less education. However, there is comparatively little evidence relating to the specific effects of questionnaire length and what little there is appears to be somewhat mixed, though in general it lends support to our hypothesis that response effects will be more likely to affect data from longer survey questionnaires.

Sample Design and Response Rates

The ESS is intended to cover individuals aged 15 and over (no upper age limit) resident within private households in each country, regardless of their nationality, citizenship or language. In the present study, resource constraints restricted us to relatively small sample sizes, but participating fieldwork agencies were instructed to use the best possible probability sample design available (in all cases, this was developed in consultation with one of the authors, who is a member of the ESS panel of sampling experts). Sample designs were allowed to vary cross-nationally, depending on the availability of sampling frames in each of the different countries. In all cases, the samples selected were of households, so at the first contact, interviewers were required to use a random selection procedure to identify a target respondent and no substitutions were allowed.

Three out of four of the countries used list-assisted methods of RDD sampling, while the remaining country, Switzerland, used the telephone directory, which, at the time of sample selection, was understood to provide a level of coverage of around 97% of resident households. Note, however, that in Switzerland the sample was restricted to the French-speaking population. In the remaining countries, the samples were intended to represent the ESS population, though in practice, were representative only of those households with fixed-line telephones. The proportion of cell-phone only households varies widely in

Europe, but was estimated to be around 6%[1] in Germany, 34% in Hungary, 22% in Poland and 1% in Switzerland (see Roberts, Eva and Widdop, 2008) at the time the research was undertaken. Given the aim of the present study was to examine relative differences between the experimental groups (to which participants were randomly assigned), rather than to make inferences to the population as a whole, the resulting under-coverage was not deemed to be overly problematic.

In each country a probability sample of phone numbers was selected and a random procedure subsequently used to assign sample members to one of the three experimental groups. Approximately 20% of the sample was allocated to group B and 40% to each of groups A and C. Table 1 shows the issued and achieved sample sizes in each country. As predicted, response rates varied across the treatment groups, and were generally higher for the shorter questionnaire (version C, part 1 only) than for the longer questionnaires.

Table 1 – Issued and achieved sample sizes by country and treatment group

	Hungary[1]	**Germany**	**Poland**	**Switzerland[2]**
Total issued sample	1000	1545	1422	859
Last call outcome:				
Contact, interview	252	369	339	342
Non-contact	101	250	270	62
Contact – no interview	78	41	140	174
Refusal	569	860	211	247
Not eligible	-	25	462	34
Number of complete interviews	210	329	292	293
Response rates:				
Version A – 60 mins (%)	18.0	20.3	32.1	37.9
Version B – 45 mins (%)	22.0	25.0	37.0	39.5
Version C (2 * 30 mins) (%)	23.5	21.3	25.6	26.8
C part 1 only (30 mins) (%)	31.5	25.2	32.4	50.2
Overall response rate[3] (complete interviews only) (%)	**21.0**	**21.6**	**30.4**	**35.5**

Notes: [1]In Hungary, the data we have about the sample refer only to the starting sample of 1000 cases. The issued samples in the other countries include ineligibles. [2]Due to limited resources the issued sample in Switzerland was smaller than in the other countries. [3]ESS response rates are calculated as the number of complete interviews divided by the eligible sample and is equivalent to AAPOR Response Rate 1. Rates reported here are based on the outcome of the last call attempt to the sampled number. Response rates are unweighted.

[1] Data from ESS round 3 (2006), edition 3.2. The ESS asks whether respondents have a fixed line telephone in their accommodation and if they personally have a mobile phone. Mobile-only households are defined here as households where there is no fixed line phone and at least one resident personally has a mobile phone (see Roberts, Eva and Widdop, 2008).

Survey agencies were permitted to conduct interviews for all treatment groups in multiple parts if requested by the respondent. In practice, only a small number of cases opted to do so. In order to explore variation in data quality as a function of questionnaire length, however, we analyse data only from those cases that completed the interviews according to the intended protocol. For group A, this meant completing the full interview in a single appointment, and for group C, this meant completing the two 30-minute interviews in two separate appointments. Around 30% of group C respondents elected to complete both interviews in a single appointment, but because of the need to differentiate respondents receiving the shortest questionnaires with respect to our module of interest (group C) from those receiving the longest questionnaire (group A), we discarded them from our analysis. Henceforth, group C refers to only those respondents that completed the second 30 minute interview in a separate appointment to the first.

Several studies have found evidence of cultural differences in response effects such as extreme response style and acquiescence (Hui and Triandis, 1989; Clarke, 2001; Villar, 2009) and social desirability bias (Johnson and van der Vijver, 2003). Nevertheless, although there is evidence to suggest measurement errors might vary cross-nationally, we had no reason to believe that the predictors of satisficing would be different in each country because satisficing behaviour is assumed to depend on fundamental and widespread psychological mechanisms. We therefore pooled the national samples to lend the maximum statistical efficiency to our analyses. However, we were aware that whatever cultural differences there might be in the tendency to favour particular response styles, there are also likely to be other factors at play, such as field agency 'house effects', translation issues and so forth. Given this, our strategy was to include country controls in our multivariate analyses. Table 2 shows the number of cases analysed in each of the three groups of interest.

Table 2 – Cases analysed by treatment group and country

	Group A (60 minutes)	**Group B (45 minutes)**	**Group C (30 minutes)**	**Total**
Germany	123	70	107	230
Hungary	69	43	42	111
Poland	99	58	75	174
Switzerland	131	80	42	173
Total	422	257	266	688

As random allocation to treatment took place before fieldwork began, there is potential confounding of non-response error with treatment group. Response propensity could be a function of a) the decision to participate in an interview of a particular *expected* duration; and b) availability and willingness to complete the interview in one (for groups A and B) or two parts (for group C). This propensity is not constant across conditions, as Table 1 showed, so there is a potential confound between treatment group and respondent characteristics.

In order to draw valid conclusions about the actual treatment, or causal effect of questionnaire length on measurement error, we need to control possible confounds. We therefore attempted to assess the extent of differential selection bias by comparing respondents between treatment groups on a range of socio-demographic and substantive variables. We chose variables most likely to be correlated with the decision to participate in surveys generally, and, specifically, in long telephone interviews. The aim was to identify observed variables potentially acting as common causes or correlates (see Groves, 2006) of nonresponse and measurement error in order to control for them in the subsequent analyses of satisficing across groups.

Questionnaires

The study used different versions of the ESS questionnaire, adapted specially for telephone administration[2] for each treatment group (see table A1 in the appendix). For group A, the questionnaire was essentially identical in terms of structure and length to the standard face-to-face questionnaire. For group B, respondents answered the same questionnaire as group A minus one of two 'rotating' modules[3] (the timing of life module was excluded). For group C, the full questionnaire was split into two parts. To ensure adequate background information about the respondent was collected in part 1 (in case he or she refused to respond to part 2), it was necessary to split the module of socio-demographic items, and move the first set of these items to the end of the first part interview, between the two rotating modules. The remainder of the socio-demographic module appeared at the end of

[2] The adaptations included deleting references to showcards and adding instructions to interviewers to read out the response options or a description of the rating scale to be used. For a small number of questions, which rely on more elaborate showcards in face-to-face mode, additional changes were necessary, such as converting the question to an open-ended format, breaking the question into two or more parts, or collapsing response categories to reduce the overall number of options to be read out by the interviewer.

[3] The ESS questionnaire comprises a core module and, typically, two rotating modules covering new substantive topics that are unique to each round (though they may in future be repeated).

part 2, as in the original questionnaire. These structural adaptations to the questionnaire meant that the placement of the module on personal and social wellbeing varied between versions. For group A, it was preceded by 141 items, for group B, by 86 items, while for group C participants whose part 2 interview took place on a separate occasion to their part 1 interview it was preceded by just 36 items, near the start of the second interview. To examine the effect of varying the length of the questionnaire (in other words, the number of preceding questions asked) on data quality, we focus on responses to this particular module. Note that in terms of overall questionnaire length, groups A and C answered the longest questionnaires, and group B answered the shortest. However, in terms of the total number of items asked in the interview prior to the module on wellbeing, it was group C who answered the shortest questionnaire, group A who answered the longest, while group B was between the two. The module contained 46 items that were applicable to all respondents, all designed to measure different dimensions of wellbeing and happiness (question wording and response options are shown in table A2 in the appendix).

Methods of Analysis

To examine the extent to which conclusions about variation in the extent of satisficing between groups can be attributed to questionnaire length, rather than to differential nonresponse across the treatment groups or other selection biases, we first explore variation between different groups of respondents in the study. The main comparison of interest is between the groups who were administered questionnaires of different lengths. However, we also compare the group C respondents we analysed with those excluded from the analysis, to assess the extent to which our decision to analyse data only from those completing the two interview parts on separate occasions resulted in a biased sample.

Our analysis of bias utilizes a number of variables from the socio-demographic module of the survey, and from the first two modules of questions (the placement of which was constant across all the treatment groups). To compare groups, we test for bivariate differences in response distributions, calculating t-tests for the equality of means on continuous variables (including 11-point rating scales) and Chi-square tests of whether responses to categorical and ordinal variables are independent of treatment group. We first examine background variables likely to be linked to response propensity – namely, respondent sex, age, education, occupation and income. We then compare responses on a selection of social, psychological and behavioural variables that could relate to willingness

to participate (e.g. measures of social trust, political interest and participation; and participation in social activities).

The main focus of our analysis is on the extent of satisficing in each of the treatment groups as indicated by item non-response, non-differentiation, acquiescence, preference for mid-points and response order effects in rating scales (primacy, recency and extremeness).

To assess item non-response between the groups, we compare the mean proportion of items in the module for which the respondent had given either a refusal, a 'Don't Know' response, or for which there was simply no recorded data (a missing value coded 'No answer')[4]. For each of the other satisficing indicators, we computed scores based on responses to a series of ten sets of questions sharing common response scales (see Appendix A2 for details). These were:

- 19 agree/disagree items (presented in four sets) using a five-point rating scale, fully labelled: *agree strongly; agree; neither agree nor disagree; disagree; disagree strongly*.
- 15 items (in a single set) using a four-point scale labelled: *none or almost none of the time; some of the time; most of the time; all or almost all of the time*.
- 4 items (in two sets) using an anchored 7-point scale, where the end-points were labelled *none of the time* and *all of the time*.
- 5 items (in one set) using an anchored 7-point scale, where the end-points were labelled *not at all* and *a great deal*.
- 4 items (in two sets) using 11-point response scales with the endpoints labelled *extremely satisfied* and *extremely dissatisfied*[5].

Scores for each indicator were calculated by counting the number of times the respondent selected a given response option in a given item set (or combination of sets) and rescaling the score to range from 0 to 1. For *acquiescence*, the score was calculated as the number of times the respondent selected the agree category on the 19 agree/disagree items. For *midpoint use*, the number of middle responses given to items with odd-numbered scales was counted. For *primacy* and *recency*, first and last response category responses were counted for two different types of question: items with ordinal response categories and items with

[4] Note that 'Don't Know' is not explicitly offered as a valid response in the ESS questionnaire, but interviewers are instructed to record all refusals and 'Don't Knows' without probing the respondent for a valid response.

[5] Question wording for all items is shown in the appendix.

rating scales with end-point labels. We make this distinction on the grounds that the mechanism underlying preference for first and last category responses may be different for each type; for the former, we might expect primacy resulting from satisficing on the telephone to account for preference for first category responses, while with the latter, we might expect a preference for end-points of a scale to be driven by extreme response style (e.g. Greenleaf, 1992), though the expected direction of the effect is not clear[6]. For *non-differentiation* the score was computed as the maximum number of times respondents selected the same response alternative for items presented in sets containing at least four items with the same response scale, regardless of which was the preferred response alternative. Note that all item sets contained a mix of positive and negative statements (see table A2 in the appendix for question wording).

Given the number of sets of items available for analysis, we were able to compare groups on multiple indicators of satisficing. To get an initial picture, we simply compared scores between groups on all available indicators (shown in table 4), using t-tests to test the difference in means. We then estimate a series of four nested OLS regression equations for summary indicators of each form of satisficing based on all available question sets to evaluate group differences while controlling for observed selection biases, as well as the country of data collection. Further details about which covariates were included in each model and how they were coded are presented in the next section.

RESULTS

Preliminary analysis

Before presenting the results of our tests of satisficing, we first present the results of our analysis of selection bias across the treatment groups. As mentioned, this analysis was aimed at ensuring that differences in satisficing could not be attributed to selection bias between the two groups resulting from differential nonresponse or from the decision to analyse data only from participants completing the survey according to the intended protocol. We present the results of statistical tests comparing the groups on sociodemographic variables (Table 3a) and on selected questionnaire variables (Table 3b).

[6] Note that some authors have combined preference for end-points and middle responses alternatives in the same indicator of satisficing (e.g. Kaminska et al, 2010). Given the ongoing debate about whether midpoint preference constitutes a form of satisficing or not, we prefer to treat each type of response effect separately.

Table 3a – Comparisons between treatment groups on selected socio-demographic variables. All countries.

Variable	Version A	Version B	A/B	Version C in 2 parts	A/C	B/C	Version C in 1 part	C in 1/ C in 2	Version C, part 1 only	All of C / C part 1 only
	n=422	N=257		n=266			n=107		n=167	
	% (SE)	% (SE)		% (SE)			% (SE)		% (SE)	
Male (%)	45.5 (.02)	45.1 (.03)		39.8 (.03)	'		36.4 (.05)		42.5 (.04)	
Mean age (years)	47.6 (.84)	47.3 (1.1)		50.8 (1.04)	*	*	48.6 (1.04)		49.2 (1.4)	
Area of residence (%)										
A big city	15.2 (.02)	16.3 (.02)		14.3 (.02)			15.9 (.04)		-	
Suburb or outskirts of city	11.6 (.02)	12.8 (.02)		15.0 (.02)			11.2 (.03)		-	
Town or small city	34.4 (.02)	30.7 (.03)		35.0 (.03)			31.8 (.05)		-	
Country village	32.9 (.02)	36.2 (.03)		31.2 (.03)			38.3 (.05)		-	
Farm or home in countryside	5.9 (.01)	3.9 (.04)		4.5 (.01)			2.8 (.02)		-	
Currently in paid work (%)	44.2 (.02)	42.4 (.03)		45.1 (.03)			47.7 (.05)		49.1 (.04)	
Main activity in last 7 days (%)										
Paid work	50.6 (.02)	50.2 (.03)		45.9 (.03)			43.0 (.05)		49.1 (.04)	
Education	10.5 (.02)	7.8 (.02)		6.8 (.02)	'		9.3 (.03)		7.2 (.02)	
Retired	24.9 (.02)	26.5 (.03)		35.3 (.03)	**	*	33.6 (.05)		29.9 (.04)	
Housework, caring for children	7.4 (.01)	7.8 (.02)		5.3 (.01)			6.5 (.02)		4.2 (.02)	
Other[1]	6.9 (.01)	4.3 (.02)		6.8 (.02)			7.5 (.03)		9.6 (.02)	
High income[2] (%)	32.6 (.03)	26.5 (.03)		25.1 (.03)	*		26.4 (.05)		24.6 (.04)	
Mean years of education	14.0 (.17)	14.0 (.23)		14.3 (.26)			14.3 (.31)		13.8 (.34)	
Household size	2.8 (.07)	2.72 (.08)		2.7 (.09)			2.7 (.14)		-	

Notes: [1]'Other' includes those who were unemployed and permanently sick or disable. [2]High income' was calculated by combining the top 3rd income groups for each country. ***p<0.001 **p<0.01 *p<0.05 'p<0.1. – No data available for nonrespondents to part 2 of version C.

Table 3b – Comparisons between treatment groups on selected questionnaire variables. All countries.

Variable	Version A	Version B	A/B	Version C in 2 parts	A/C	B/C	Version C in 1 part	C in 1/ C in 2	Version C, part 1 only	All of C / C part 1 only
	n=422	n=257		n=266			n=107		n=167	
	% (SE)	% (SE)		% (SE)			% (SE)		% (SE)	
How satisfied are you with your life (mean on 11-pt scale)	7.0 (.12)	6.7 (.14)		6.7 (.15)			6.5 (.25)		6.9 (.18)	
General health (% Very good or good)	66.8 (.02)	63.4 (.03)		53.4 (.03)	***	*	56.1 (.05)		-	
General health (% Very bad or bad)	7.1 (.01)	7.4 (.02)		6.7 (.02)			11.2 (.03)		-	
Hampered by long-standing illness (% A lot or Some)	26.1 (.02)	27.2 (.03)		38.3 (.03)	***	**	27.1 (.04)	*	-	
Participation in social activities (% Less or much less than most)	33.8 (.02)	35.0 (.03)		40.5 (.03)	*		46.2 (.05)		-	
Participation in social activities (% more or much more than most)	21.9 (.02)	18.9 (.03)		15.2 (.02)	*		16.0 (.04)		-	
Most people can be trusted (mean on 11-pt scale)	5.0 (.13)	5.0 (.15)		5.0 (.15)			4.7 (.25)		4.5 (.20)	'
Most people would try to be fair (mean on 11-pt scale)	5.9 (.12)	6.1 (.14)		5.9 (.14)			5.6 (.28)	'	5.5 (.21)	
People mostly try to be helpful (mean on 11-pt scale)	4.9 (.11)	5.1 (.14)		4.8 (.14)			4.8 (.24)		4.4 (.20)	'
Interest in politics (% Very interested)	17.8 (.02)	13.6 (.02)	'	15.8 (.02)			12.1 (.03)		13.2 (.03)	
Interest in politics (% Not at all interested)	11.4 (.02)	7.8 (.02)	'	6.8 (.02)	*		12.1 (.03)		9.6 (.02)	
Politics so complicated you can't understand (% Never)	12.1 (.02)	8.7 (.02)	'	8.3 (.02)	'		6.7 (.02)		8.0 (.02)	
Politics so complicated you can't understand (% Frequently)	5.7 (.01)	9.4 (.02)	*	9.4 (.02)	*		12.4 (.03)		8.6 (.02)	
Difficult or easy to make up mind about politics (% difficult or v difficult)	26.1 (.02)	27.3 (.03)		23.8 (.03)			20.0 (.04)		34.6 (.04)	*
Difficult or easy to make up mind about politics (% easy or v easy)	39.5 (.02)	33.2 (.03)	'	40.4 (.03)		'	42.9 (.05)		34.0 (.04)	
Voted in last election (% yes of eligible)	67.3 (.02)	74.7 (.03)	*	75.1 (.03)	*		81.3 (.04)		66.0 (.04)	*
Contacted a politician (% yes)	18.7 (.02)	22.6 (.03)		16.2 (.02)		*	19.6 (.04)		12.0 (.03)	
Worked in a political party (% yes)	6.4 (.01)	7.4 (.02)		6.4 (.02)			1.9 (.01)		4.8 (.02)	
Worked in another organisation (% yes)	19.7 (.02)	21.8 (.03)		20.0 (.03)			10.3 (.03)	*	18.0 (.03)	
Wore a campaign badge (% yes)	10.9 (.02)	9.4 (.02)		6.4 (.02)	*		10.3 (.03)		4.8 (.02)	
Signed a petition (% yes)	41.2 (.02)	40.2 (.03)		27.4 (.03)	***	**	24.5 (.04)		33.1 (.04)	
Took part in a public demonstration (% yes)	8.3 (.01)	13.6 (.02)	*	10.9 (.02)			7.5 (.03)		7.2 (.02)	
Boycotted certain products (% yes)	29.3 (.02)	28.0 (.03)		24.2 (.03)	'		26.4 (.04)		25.9 (.03)	
Mean political participation score (sum of all actions)	1.3 (.07)	1.4 (.10)		1.1 (.09)	*	*	1.00 (.12)		1.1 (.09)	

Notes: ***p<0.001 **p<0.01 *p<0.05 'p<0.1. – No data available for nonrespondents to part 2 of version C.

Respondents in the shortest questionnaire group (C) were statistically different from respondents in the longest questionnaire group (A) on three of the socio-demographic variables tested. Respondents from group C were significantly older than respondents from group A (50.8 years in C compared with 47.6 in group A; $p<.05$). They were more likely to be retired (35.3% in group C compared with 24.9% in group A; $p<.01$), and were less likely to be in the highest income category (25.1% in group C compared with 32.6% in group A). Group C respondents were also less likely to be male (39.8% compared with 45.5% in group A; $p<0.1$). Similar differences on sex, age and main activity were observed between groups B and C, but the difference on income was not observed (table 3a). In the top half of table 3b, we see that consistent with the differences in age and main activity, respondents in group C were also less likely to report very good or good health (53.4% in group C compared with 66.8% in group A; $p<.001$), were more likely to report being hampered a lot or some by long-standing illness (38.3% in group C compared with 26.1% in group A; $p<.001$) and were less likely to participate in social activities (40.5% in C reported participating less or much less than most, compared with 33.8% in group A; $p<.05$). Similar differences were observed between groups B and C and on the health variables, they were statistically significant. There were no differences between groups A and B on these variables.

In the lower half of table 3b, we see that respondents in the shortest questionnaire group also differed from respondents in the longest questionnaire group on a number of political interest and participation variables. Group C respondents were generally less likely to participate in a range of political activities compared to respondents in both groups A and B (reporting a mean of 1.1 out of a maximum of 7 types of political activity, compared with 1.3 (group A) and 1.4 (group B); $p<.05$). Group C respondents were less likely to report being 'not at all interested' in politics (6.8% compared with 11.4% in group A; $p<.05$) but were more likely to report frequently finding politics 'so complicated you can't understand it' (9.4% in group C compared with 5.7% in group A; $p<.05$). However, group C respondents were more likely to report having voted in the last election (75.1% compared with 67.3% in group A; $p<.05$). The differences between groups B and C on these variables were not statistically significant; group B respondents were more similar to group C on these two latter variables and differed significantly from group A on both.

To establish whether the group C respondents who completed the interview in two separate appointments differed either from those who completed the interview in one go or those who refused to respond to part 2, we compared all three group C samples on the same set of variables. There were no statistically significant differences in the socio-demographic composition of the different types of group C respondents. However, some significant differences were observed on a few of the questionnaire variables. Compared to group C respondents finishing the complete interview (whether in one or two interviews), non-respondents to part 2 of the interview were significantly more likely to report that they found it 'difficult or very difficult' to make their mind up about politics and were less likely to report having voted in the last election. Respondents completing the whole questionnaire in two interviews were less likely than those who completed the whole interview in one interview to report being hampered by a long-standing illness and were more likely to report having worked in a political organisation (other than a political party).

The results of this initial examination persuaded us that we should control for all of the variables where we observed a statistically significant difference across conditions. As mentioned earlier, we also controlled for country of data collection. This strategy has few disadvantages, even if the bias problem turned out to be of no consequence. In fact, by including covariates of this kind, we may even increase the efficiency of the treatment estimator (Pocock et al., 2002).

Satisficing and interview length

To get an initial idea of the extent and direction of differences in satisficing as a function of questionnaire length, we first examined item non-response rates and second, conducted t-tests to compare mean scores on each of the other indicators across the three treatment groups. Overall rates of missing values on the 46 items in the well-being module were very low. Refusals and 'No answers' affected only 1 or 2 questions. The mean rate of Don't Know responses (i.e. the total number of 'Don't Know' responses divided by the number of items in the module) was also very low at 0.25 in group A, 0.26 in group B and just 0.18 in group C. If Don't Know reporting is indicative of reduced data quality, then this difference is in the expected direction – i.e. respondents were less likely to answer 'Don't Know' in the short questionnaire group – but it was not statistically significant.

Table 4 – Mean scores on satisficing indicators

Indicator	No. of Items	Mean A	Mean B	Mean C	Sig. A/B	Sig. A/C	Sig. B/C
Non-differentiation							
1) Agree/Disagree set 1 (5-pt scales)	**4**	.48	.47	.47			
2) Agree/Disagree set 2 (5-pt scales)	**8**	.36	.37	.39			
3) Agree/Disagree set 3 (5-pt scales)	**6**	.32	.30	.31			
4) Agree/Disagree set 4 (5-pt scales)	**1**	-	-	-	-	-	-
5) None or almost none/ all of the time set 1 (4-pt scales)	**15**	.32	.32	.32			
6) None of the time/ all of the time set 2 (7-pt scales)	**2**	.54	.52	.54			
7) None of the time/ all of the time set 3 (7-pt scales)	**2**	.34	.35	.33			
8) Not at all/ a great deal set 1 (7-pt scales)	**5**	.37	.37	.35			
9) Extremely satisfied/ dissatisfied set 1 (11-pt scales)	**2**	.39	.35	.34			
10) Extremely satisfied/ dissatisfied set 2 (11-pt scales)	**2**	.34	.35	.33			
Non-differentiation summary (1,2,3,5,6,7,8)	**40**	**.45**	**.44**	**.45**			
Acquiescence							
Acquiescence (1,2,3,4 - agree responses only)	19	.41	.41	.44		*	*
Use of midpoint							
Use of midpoint (Agree/Disagree sets)	19	.21	.22	.25		**	‘
Use of midpoint (7-pt scales)	9	.17	.16	.20		*	**
Use of midpoint (11-pt scales)	4	.10	.11	.11			
Use of midpoint summary (All odd-numbered scales)	**32**	**.23**	**.24**	**.28**		***	*
First category response (Primacy and Extremeness)							
Primacy (1,2,3,4 - Agree/Disagree sets)*	19	.25	.25	.20		**	**
Primacy (5 - None or almost none of the time)	15	.45	.43	.41		*	
Extremeness (9,10 - Extremely dissatisfied)	4	.02	.01	.01		‘	
Extremeness (6,7 - None of the time)	4	.14	.10	.11	*	*	
Extremeness (8 - Not at all)*	5	.13	.12	.10		*	
Extremeness (All 7-pt scales)	9	.17	.14	.13		*	
Primacy summary (ordinal scales)	**34**	**.39**	**.38**	**.34**		***	**
Extremeness summary (7- and 11-pt scales)	**13**	**.14**	**.11**	**.10**	*	**	
Last category responses (Recency and Extremeness)							
Recency (1,2,3,4 - Disagree strongly)*	19	.13	.13	.10		**	*
Recency (5 - All or almost all of the time)	15	.18	.19	.17			
Extremeness (9,10 - Extremely satisfied)	4	.10	.07	.06	*	*	
Extremeness (6,7 -All of the time)	4	.17	.17	.15			
Extremeness (8 - A great deal)*	5	.17	.15	.13		*	
Extrememeness (All 7pt scales)	9	.22	.20	.18		*	
Recency summary (ordinal scales)	**34**	**.25**	**.25**	**.21**		*	*
Extremeness summary (7- and 11-pt scales)	**13**	**.19**	**.17**	**.15**		**	

Note: ***p<0.001; **p<0.01 *p<0.05 ‘p<0.1

Consistent with our hypothesis, respondents in the longest questionnaire group (group A) were significantly more likely than respondents in the shortest questionnaire group (group C) to select the first and last category responses on both types of rating scale (see table 4). Primacy was more common on items with ordinal scales, while recency was slightly more common on rating scales with end-point labels. Group C also differed significantly from group B on the primacy and recency indicators for items with ordinal scales, whereas these groups had more similar scores for extremeness on rating scales with end-point labels (correspondingly, group B differed significantly from group A on three of the extremeness indicators). By contrast, group C respondents were more likely than respondents in either of the longer questionnaire groups to give acquiescent responses and to favour the middle response alternative. The latter was true for three of the four item sets, as well as the summary indicator. There were no statistically significant differences between any of the groups for any of the non-differentiation indicators.

The differences observed on each of the summary satisficing indicators were analysed further by estimating a series of nested OLS regression equations, the results of which are shown in tables 5a and 5b. Given that group B respondents more closely resembled the group A respondents in the bivariate analyses, we focused the multivariate analysis on the respondents from the longest (group A) and shortest (group C) questionnaire groups (interview length in the models is represented by a dummy variable where 1 indicates that the respondent was in group C).

In the first set of models (not shown in tables), we included the indicator for interview length and the controls for the country of data collection (Hungary, Poland and Switzerland), leaving Germany as the reference country as it had the largest sample size. This identified a number of significant differences between countries in the level and nature of satisficing. For example, compared to the German sample, respondents in Switzerland were less likely to differentiate scale points (i.e. they exhibited more non-differentiation), more likely to give acquiescent responses, and were more likely to select the first- or last-category response on both types of response scale. Polish respondents were also less likely than German respondents to differentiate, and more likely to acquiesce, select the midpoint and select first-category responses, while Hungarian respondents were more likely to acquiesce and give extreme responses to rating scales with end-point labels. Regarding the effect of questionnaire length on satisficing, we find that when controlling for country of

data collection, the relations observed in our bivariate analyses persist for midpoints, primacy on both types of scale, and recency on rating scales with end-point labels, but not for acquiescence and recency on ordinal categorical variables.

Table 5a Coefficients from OLS regression equations predicting nondifferentiation, acquiescence and scale midpoint use

	Nondifferentiation				Acquiescence				Midpoints			
	3		**4**		**3**		**4**		**3**		**4**	
Constant	0.388	***	0.392	***	0.393	***	0.416	***	0.356	***	0.396	***
Questionnaire length (short)	-0.011		-0.012		0.015		0.013		0.040	**	0.032	*
Country:												
Hungary	-0.013		-0.020		-0.076	**	-0.090	***	-0.016		-0.031	
Poland	0.051	**	0.050	**	0.053	**	0.048	*	-0.041	*	-0.042	*
Switzerland	-0.044	**	-0.035	*	-0.065	**	-0.059	**	-0.114	***	-0.097	***
Demographics:												
Male	0.023	'	0.018		0.014	*	0.012		-0.014		-0.010	
Age in years	0.008	***	0.008	***	0.005	*	0.005	'	-0.001		-0.001	
Age squared	-0.008	***	-0.008	***	-0.005	*	-0.005	'				
Retired	0.026		0.025		0.032		0.035		-0.061	**	-0.063	**
High income	-0.019		-0.017		0.004		0.014		-0.018		-0.009	
Years of education	-0.007	***	-0.007	***	-0.006	*	-0.005	*	-0.001		0.000	
Questionnaire variables:												
Good health			0.002				-0.010				-0.035	*
Hampered by long-standing illness			-0.010				-0.008				0.009	
Participates more in social activities			-0.021				-0.036	'			-0.046	*
Not at all interested in politics			-0.021				0.002				-0.048	'
Politics frequently too complicated			0.023				-0.012				-0.002	
Voted in last election			0.009				0.000				0.000	
Political participation score			-0.007				-0.012	*			-0.009	'
N	578		570		578		570		578		570	
R Square	0.140		0.148		0.119		0.136		0.132		0.168	

Notes: ***p<0.001 **p<0.01 *p<0.05 'p<0.1.

Table 5b Coefficients from OLS regression equations predicting preference for the first and last response categories on ordinal measures and rating scales

	Primacy on ordinal measures				Primacy on rating scales			
	3		**4**		**3**		**4**	
Constant	0.301	***	0.246	***	0.018		0.023	
Questionnaire length (short)	-0.039	*	-0.026	'	-0.046	***	-0.042	**
Country:								
Hungary	-0.022		0.001		0.200	***	0.205	***
Poland	-0.025		-0.018		0.082	***	0.077	***
Switzerland	0.072	***	0.045	*	0.059	***	0.046	**
Demographics:								
Sex	0.056	***	0.050	**	0.012		0.013	
Age	0.001		0.001		0.002	***	0.002	***
Age square								
Retired	-0.021		-0.017		-0.011		-0.008	
High income	0.022		0.005		-0.021		-0.021	
Years of education	0.002		0.001		-0.002		-0.002	
Questionnaire variables:								
Good health			0.058	**			0.007	
Hampered by long-standing illness			-0.053	**			0.014	
Participates more in social activities			0.071	***			0.010	
Not at all interested in politics			0.042				0.013	
Politics frequently too complicated			-0.012				0.012	
Voted in last election			0.004				-0.034	*
Political participation score			0.010	'			0.003	
N	578		570		578		570	
R Square	0.105		0.197		0.221		0.233	
	Recency on ordinal measures				**Recency on rating scales**			
	3		**4**		**3**		**4**	
Constant	0.129	*	0.037		0.115	**	0.081	'
Questionnaire length (short)	-0.031	'	-0.026		-0.042	**	-0.036	*
Country:								
Hungary	0.045		0.075	*	0.158	***	0.168	***
Poland	-0.005		0.000		0.014		0.013	
Switzerland	0.082	***	0.055	*	0.044	*	0.029	
Demographics:								
Sex	0.031	'	0.029	'	-0.007		-0.013	
Age	0.002	*	0.002	**	0.001	*	0.001	*
Age square								
Retired	-0.008		-0.007		-0.018		-0.013	
High income	-0.019		-0.037	'	0.006		-0.001	
Years of education	0.000		-0.001		-0.001		-0.002	
Questionnaire variables:								
Good health			0.079	***			0.045	*
Hampered by long-standing illness			0.034				-0.013	
Participates more in social activities			0.056	*			0.028	
Not at all interested in politics			0.047				0.003	
Politics frequently too complicated			0.044				0.043	
Voted in last election			-0.006				0.015	
Political participation score			0.020	**			0.002	
N	578		570		578		570	
R Square	0.064		0.126		0.106		0.134	

Notes: ***p<0.001 **p<0.01 *p<0.05 'p<0.

In the second set of equations we included a series of socio-demographic covariates to control for differences observed in the sample composition of the groups: respondent sex (coded 1 if male), age in years, age squared (age*age*.01, excluded if not statistically significant), main activity (coded 1 if retired), and income (coded 1 if in the top 3 income categories). In addition, we included the number of years of education the respondent had received to control for the commonly-observed association between education and satisficing. The regression coefficients for these models are shown in the columns labelled '2' in tables 5a and 5b. This step of the analysis revealed some significant associations between certain socio-demographic characteristics and different response styles, but controlling for sample differences did not affect the relationships observed previously between questionnaire length and satisficing: respondents in the longest questionnaire group were still more likely to use first- and last-category responses, while those in the shortest questionnaire group preferred the midpoint. While questionnaire length did not predict non-differentiation or acquiescence, the number of years of education and age squared were both negatively and significantly associated with both, consistent with previous studies of satisficing. Age was positively and significantly associated with primacy extremeness on rating scales and both forms of recency, whereas being retired was negatively and significantly associated with midpoint use.

In the final set of equations we included a number of questionnaire variables where we had observed differences between the samples. These included: being in good health (coded 1 if respondent reported 'very good or good health'); being hampered by a long-standing illness (coded 1 if respondent reported being hampered 'a lot or some'); participation in social activities (coded 1 if 'more or a lot more than most'); interest in politics (coded 1 if 'not at all interested'); finding politics so complicated you can't understand it (coded 1 if 'frequently'); voting in the last general election (coded 1 if voted); and an index of political participation based on reported participation in seven activities (ranging from 0 to 7). The equation coefficients are shown in tables 5a and 5b in the columns labelled '3'. Controlling for selection bias on these questionnaire variables had the effect of weakening the association between questionnaire length and satisficing for all the variables where there was still a significant relation in model 2,

but did not account for it completely (though the relation between questionnaire length and primacy on ordinal measures only approached significance after including all the covariates). Respondents in group C were still significantly more likely to select the midpoint on odd-numbered scales (table 5a), while respondents in group A were significantly more likely to favour the first and last category of rating scales with end-point labels.

DISCUSSION

Survey designers frequently find themselves under pressure to add new items to already long survey questionnaires. While the addition of extra questions may seem cost effective in the context of a large-scale face-to-face data collection exercise, the true costs in terms of data quality may be underestimated. Potential respondents may be less inclined to participate in a survey expected to last a long time, and those who do agree to take part, may find their motivation to respond in a thoughtful way decreases over the course of a long interview; each process resulting in a net increase in total survey error. Long questionnaires designed for face-to-face administration present a more specific challenge in the current survey climate, where the rising costs of conducting fieldwork in person and of maintaining acceptable response rates have led survey designers working on repeated studies (such as the European Social Survey) to consider either switching to alternative, cheaper data collection modes, or employing a mix of modes either to reduce costs, or minimise selection biases (from non-coverage or non-participation) associated with the main alternatives to face-to-face (notably, telephone and web surveys). Common practice is based on the assumption that long questionnaires are unsuitable for administration in modes other than face-to-face interviews, yet there is still surprisingly little empirical evidence concerning the actual effects of questionnaire length on survey errors in different data collection modes. In this article, we investigated the impact of questionnaire length in a telephone survey on a range of response effects commonly observed in survey data, which, according to the theory of satisficing (Krosnick, 1991), are more likely to occur under conditions where respondent' ability and motivation to respond carefully are reduced and the difficulty of the response task is increased. Increasing questionnaire length was hypothesised to

contribute to such conditions, leading to the prediction that respondents will be more likely to satisfice on items placed later in the questionnaire.

Consistent with this prediction, the results of our investigation showed that respondents answering longer questionnaires by telephone were more likely to exhibit response order effects in their answers, a response strategy that has been described as an example of weak satisficing by Krosnick (1991). Specifically, they were more likely to select the first-category response on fully labelled ordinal scales and more likely to give extreme, last-category responses on rating scales with end-point labels. This finding is consistent with those of other studies that have found response order effects to be more likely under conditions that foster satisficing (Krosnick, 1991), and provides new evidence that questionnaire length – at least in a telephone interview - can contribute to such conditions. The fact that primacy and recency effects varied as a function of response format may help to explain why previous studies of order effects on rating scales have found a mix of primacy and recency (Krosnick, Judd and Wittenbrink, 2005) and underlines the importance of differentiating scale type when studying this form of satisficing. According to our results, fully-labelled ordinal scales appear to function in a similar way to lists of unordered categories in a telephone interview, giving rise to primacy effects as respondents select the first answer that corresponds to their attitude (consistent with the findings of Kalton, Collins and Brook, 1978). Such an effect might also be the result of satisficing respondents wishing to interrupt the interviewer as he/she reads out the full list of response categories. By contrast, we observed recency effects in the form of extreme responses to rating scales with end-point labels, probably because this response was freshest in memory once the interviewer finished reading the question and describing the scale.

While respondents in the longest questionnaire group were more likely to respond at the endpoints of rating scales, the respondents in the shortest questionnaire group were significantly more likely to select the middle alternative on odd-numbered rating scales. This finding runs counter to what we had expected: we predicted we would see more evidence of each form of satisficing we investigated among respondents who answered the longest questionnaire. Yet this hypothesis was underpinned by the

assumption that all our measures were indicative of satisficing, which may not be true in the case of preference for midpoints. In fact, this remains the subject of much debate in the research literature on questionnaire design. For example, O'Muircheartaigh, Krosnick and Helic (2000) found that data quality was higher when the midpoint was offered than when it was not, which would not be expected if the midpoint were attracting satisficers. They also found no clear relation between education and midpoint selection. These findings reinforce those of previous studies, which have shown that midpoint respondents tend, rather, to have higher levels of education (Narayan and Krosnick, 1996; Krosnick, Narayan, and Smith, 1997), and tally with the conclusions of a recent study by Kulas and Stachowski (2009), which suggests that selecting the middle alternative may actually be *more* cognitively demanding for respondents and, therefore, less 'easy-to-select' and 'easy-to-defend' by satisficers than has been suggested elsewhere (see Krosnick, Judd and Wittenbrink, 2005). If this is right, then we perhaps ought not conclude that our findings falsify the hypothesis that longer questionnaires encourage satisficing. Rather, we consider alternative explanations as to why we found greater preference for midpoints among respondents in the shortest questionnaire group.

One possibility is that midpoint responding arises not from satisficing, but rather from respondents' concerns to present themselves to interviewers in socially desirable ways (Sturgis, Roberts and Smith, 2010). Aside from the number of preceding questions asked and other known predictors of satisficing (namely, education), we assumed that the most likely causes of differences in response quality would be differences in the samples resulting from our decision to only analyse data from specific subsets of respondents. An alternative explanation, however, is that for respondents in Group C, participation in a previous interview per se may have influenced how they answered questions in their second interview, leading them, in particular, to give less extreme and more neutral, socially acceptable responses. Further support for the interpretation that the social context of the second interview led respondents to answer questions in a more socially polite way, comes from our bivariate analyses, which showed that respondents in the shorter questionnaire group were also more inclined to give acquiescent responses (i.e. to agree with agree-disagree scales). While numerous

studies have found acquiescence to be more common under conditions that encourage satisficing (see Krosnick and Fabrigar, forthcoming, for a review), acquiescence has also been attributed to social desirability motivations, in particular, the desire to defer, out of politeness, to the position assumed to be held by the interviewer reading out the assertion in the question (Saris et al., 2009). In our multivariate models, this effect dropped out once we controlled for the country of data collection, but we cannot exclude the possibility that in some countries at least, respondents in the shorter questionnaire group gave more acquiescent responses to those in the longer questionnaire group because the second interview constituted a different social context to that of the first. Indeed, such an interaction between country and questionnaire length is not implausible given that cross-national variations in survey climate may make respondents in countries where telephone interviewing is widely used more tolerant of long interviews (Roberts, Eva and Widdop, 2007).

Although there may be some suggestion in our data of differential rates of acquiescence by questionnaire length in some of the countries where we conducted the experiment, overall, and consistent with the findings of other studies (e.g. Clancy and Wachsler, 1971), our analysis found no relationship between questionnaire length and this form of satisficing. Similarly, we found no evidence that rates of nondifferentiation between scale points varied with the number of questions asked. We predicted that the design of the question module we analysed, which consisted of long batteries of items using the same response options, would make it particularly susceptible to this form of satisficing, which has been found in other studies to be more likely to occur on items placed later in the questionnaire (e.g. Herzog and Bachman, 1981; Kraut et al., 1975). Even before controlling for country and other variables, we found no evidence that the number of prior questions asked influenced the likelihood of selecting the same response option for every item in a battery. It is noteworthy that all of the sets of items for which we computed nondifferentiation scores contained a mix of positively and negatively worded items, thus providing a more conservative test of our hypothesis, and, reassuringly, there were no respondents who used the same scale-point for every item in a set. Nevertheless, while we found no effect of questionnaire length on rates of nondifferentiation, we did find – consistent with other

research on satisficing – that education was a significant predictor of both nondifferentiation and acquiescence, with higher rates of satisficing among respondents with fewer years of education. On this basis, we cannot conclude that respondents did not adopt these types of response strategy, but rather that questionnaire length was not an important mechanism underlying the effects. In other words, in the case of nondifferentiation and acquiescence, respondents with lower education appear to be more likely to satisfice, irrespective of the length of the survey questionnaire.

In summary, our findings provide only partial support for the hypothesis that respondents are more likely to satisfice when answering long survey questionnaires than they are when answering shorter questionnaires. While we find evidence to suggest that respondents to longer questionnaires are more likely to exhibit response order effects in their data, we find no support for the hypothesis that questionnaire length significantly influences the likelihood or extent of nondifferentiation and acquiescence. At the same time, our finding that midpoints were selected more frequently by respondents to the shortest questionnaire suggests that other mechanisms – notably, social desirability concerns – may have played a part in influencing respondents answers. This, in turn, highlights the need to investigate other plausible explanations, in addition to satisficing, for response effects frequently observed in survey data. This is particularly important when assessing the quality of telephone survey data and when making decisions about the optimal length of a telephone survey questionnaire. Our findings point to difficulties with both the design solutions we considered for how to adapt the ESS face-to-face questionnaire to make it suitable for telephone administration. On the one hand, attempting to administer the full-length questionnaire, or reducing it in length by only one module, led to satisficing among certain respondents. On the other, dividing the full questionnaire in two and administering it in separate interviews, appeared to introduce other types of response effect, perhaps as a result of altering the social context of the interview. Both findings lend further support to the received wisdom that questionnaire length should be kept to a minimum, not only to avoid break-offs and increased measurement error from satisficing, but additionally, to minimise the need to administer the questionnaire over

multiple interview appointments; a not uncommon practice, the effects of which on data quality have not previously been considered.

Certain limitations of the design of our study suggest some caveats to our conclusions. While our analysis attempted to control for the presence of selection bias resulting from the decision to analyse data from a subsample of our respondents, we were not able to control for the possibility that the context in which the module of questions on well-being were presented may have been affected by our modifications to the overall order of the questionnaire (including cutting altogether a module out of version B). For group A, the module on well-being was preceded by a module of questions on the timing of life events, while for groups B and C, it was preceded by core questions on religious practice and ethnic and national identity. Given the ample research evidence documenting context effects in surveys (e.g. Bradburn and Mason, 1964; Tourangeau and Raskinski, 1988) we cannot exclude the possibility that question order may partially account for the differences we observed between groups A and C. However, the presence of differences between groups B and C (for whom the context was the same) should give some reassurance that context was less influential than other factors. Equally, it seems unlikely that context effects resulting from questions asked immediately prior to the start of the module we analysed would have extended beyond the first few items. Indeed, a perhaps greater concern is that once in the flow of answering questions about their wellbeing, respondents may have found the topic relatively involving compared to other parts of the questionnaire, thereby reducing the likelihood of satisficing, and perhaps diminishing the hypothesised effects of questionnaire length. Nevertheless, we acknowledge that the presence of confounds from question order and the fact that group C's answers to the module were recorded in a second interview, underlines the need for more purposely-designed studies in future to investigate further the effects of questionnaire length and follow-up interviewing on response quality in telephone interviews.

We decided to pool data from each of the four countries that participated in our study on the grounds that although we might expect differences across countries on the response effects we investigated, we had no reason to assume that the predictors of

satisficing would vary cross-nationally. In fact, we found cross-national differences on all of the satisficing indicators, and for half of them (nondifferentiation, acquiescence and primacy and recency on ordinal categorical measures), the effect of country of data collection was greater than the effect of questionnaire length. While our aim was not to attempt to unravel these effects here, it is likely that a deeper exploration of the different predictors of satisficing in each country – were sample size to permit such analysis – could shed more light on our overall findings. In particular, the cross-national differences we observed lend further weight to our conclusion that other influences on data quality, including the social and cultural dynamics of the interview setting itself, should be taken into account alongside the hypothesised predictors of survey satisficing.

REFERENCES

Bradburn, N.M. and Mason, W.M. (1964). The effect of question order on responses. *Journal of Marketing Research*, vol. 1, 57-61.

Cannell, C., Miller, P. and Oksenberg, L. (1981) Research on interviewing techniques. In S. Leihardt (Ed.), *Sociological methodology* (pp. 389-437). San Francisco: Jossey-Bass.

Carp, F.M. (1974). Position effects on interview responses. *Journal of Gerontology*, 29, 581-587

Clancy, K.J. and Wachsler, R.A. (1971) Positional effects in shared-cost surveys. *Public Opinion Quarterly*, 35, 258-265.

Clarke III, I. (2001) Extreme response style in cross-cultural research. *International Marketing Review*, vol. 18, no. 3, 301-324.

de Leeuw, E. (1992). *Data quality in mail, telephone, and face-to-face surveys*. Amsterdam: TT Publications.

de Leeuw, E., & van der Zouwen, J. (1988). Data quality in telephone and face-to-face surveys: A comparative analysis. In R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283-299). New York: Wiley.

Ferber, R. (1966), Item Nonresponse in a Consumer Survey, *Public Opinion Quarterly*, 30, pp. 399-415.

Galesic, M. and Bosnjak, M. (2009) Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opinion Quarterly, *73 (2): 349-360.*

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56(3): 328-351

Groves, R. M. (1979). Actors and Questions in Telephone and Personal Interview Surveys. *Public Opinion Quarterly, 43*(2), 190-205.

Groves, R.M. (2006). Nonresponse Rates and Nonresponse Bias in Household Surveys. *Public Opinion Quarterly* 70, 646-675.

Herzog, A.R. and Bachman, J.G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, 45, 549-559

Holbrook, A. L., Cho, Y. I., & Johnson, T. P. (2006). *Extreme response style: Style or substance.* Paper presented at the Annual meeting of the Americal Association for Public Opinion Research, Montreal, Canada.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone vs. Face-to-Face Interviewing of National Probability Samples With Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias. *Public Opinion Quarterly, 67*, 79-125.

Hui, H.C. and Triandis, H.C. (1989) Effects of culture and response format on extreme response style. Journal of Cross-Cultural Psychology, vol. 20, no. 3, 296-309

Jabine, T., Straf, M., Tanur, J. and Tourangeau, R. (1984). *Cognitive aspects of survey methodology: building a bridge between disciplines.* Washington, DC: National Academy Press.

Jäckle, A., Roberts, C. E., and Lynn, P. (2006). Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes (Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project). *ISER Working Paper,* 2006-41. Colchester: University of Essex.

Johnson, T.P. and Van de Vijver, F.J.R. (2003). Social desirability in cross-cultural research. In J.A. Harkness, F.J.R. Van de Vijver and P.Ph. Mohler (Eds.) *Cross-cultural survey methods*. Hoboken, NJ: Wiley

Kalton, G., Collins, M. and Brook, L. (1978). Experiments in wording opinion questions. *Applied Statistics,* 27, 149-161.

Kaminska, O., McCutcheon, A.L., and Billiet, J. (2010, forthcoming). Satisficing among reluctant respondents in a cross-national conference. *Public Opinion Quarterly*.

Klopfer, F.J. and Madden, T.M. (1980). The middlemost choice on attitude items: Ambivalence, neutrality, or uncertainty. *Personality and Social Psychology Bulletin,* 6, 97-101

Kraut, A.I., Wolfson, A.D. and Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology*, 60, 774-776.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236.

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly, 51,* 201-219.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50,* 537-567.

Krosnick, J.A. (2002). The causes of no-opinion responses to attitude measures in surveys: They rarely are what they appear to be. In R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (Eds.). *Survey Nonresponse*. New York: Wiley, 88-100.

Krosnick, J. A., & Fabrigar, L. R. (Forthcoming). *The handbook of questionnaire design*. New York: Oxford University Press.

Krosnick, J.A., Judd, C.M., and Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B.T. Johnson and M.P. Zanna (Eds.). *The handbook of attitudes*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In M. T. Braverman & J. K. Slater (Eds.), *Advances in Survey Research* (Vol. 70, pp. 29-44). San Francisco: Jossey-Bass Publishers.

Kulas, J.T., and Stachowski, A.A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality,* 43(3), 489-493

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

O'Muircheartaigh, C., Krosnick, J.A. and Helic, A. (2000) *Middle alternatives, acquiescence and the quality of questionnaire data*. The Harris School Working Paper Series, 01.03. http://www.harrisschool.uchicago.edu/faculty/web-pages/colm-omuircheartaigh.asp

Pocock, S.J., Assmann, S.E., Enos, L.E, Kasten, L.E. (2002). *Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems*. Statistical Medicine. 21(19), pp 2917-30.

Roberts, C., Eva, G., and Widdop, S. 2008. *Assessing the Demand and Capacity for Mixing Modes of Data Collection on the European Social Survey: Final Report of the Mapping Exercise*. Unpublished manuscript, available on request from the Centre for Comparative Social Surveys, City University, London.

Roberts, C., Jäckle, A. and Lynn, P. (2006) Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-Face and Telephone Surveys. *Proceedings of the Survey Research Methods Section*, Washington DC: American Statistical Association.

Saris, W.E., Revilla, M., Krosnick, J.A., and Schaeffer, E. M. (2009) *Comparing Questions with Agree/Disagree Response Options to Questions with Item-Specific Response Options*. RECSM Working Paper 1.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic.

Simon, H.A. (1956). *Rational choice and the structure of the environment*. Psychological Review, Vol. 63 No. 2, 129-138.

Sturgis, P., Roberts, C. and Smith, P. (2010). *Middle alternatives revisited: How the neither/nor response acts as a 'face-saving' way of saying 'I don't know'*. Southampton Statistical Sciences Research Institute, Working Paper M10/01, http://eprints.soton.ac.uk/73620/2/s3ri-workingpaper-M10-01.pdf

Tourangeau, R. and Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, vol. 103, no. 3, 299-314.

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

Villar, A. (2009). *Agreement answer scale design for multilingual surveys: Effects of translation-related changes in verbal labels on response styles and response distributions.* PhD dissertation. Lincoln, Nebraska: University of Nebraska.

Appendix

Table A1 – Questionnaire structure by version

Section		*# of Items*	*Topics*	*Version A*	*Version B*	*Version C Part 1*	*Version C Part 2*
A	Core	10	Media; social trust	1	1	1	-
B	Core	40	Politics, including: political interest, efficacy, trust, electoral and other forms of participation, party allegiance, socio-political orientations	2	2	2	-
C	Core	36	Subjective well-being, social exclusion; religion; perceived discrimination; national and ethnic identity	3	3	-	1
D	Rotating module	55	Timing of life; the life course; timing of key life events, attitudes to ideal age, youngest age and oldest age of life events, planning for retirement	4	-	3	-
E	Rotating module	55	Personal and social well-being, helping others, feelings in the last week, life satisfaction, satisfaction with work.	**5**	**4**	-	**2**
F	Core	73	Socio-demographic profile, including: household composition, sex, age, type of area, education & occupation of respondent, partner, parents, union membership, income, marital status	6	5	-	-
X	Core	25	Socio-demographic profile, part 1- including: sex, age, education & occupation of respondent, income, marital status	-	-	4	-
Y	Core	56[1]	Socio-demographic profile, part 2 - including: household composition, type of area, education & occupation of respondent's partner, parents, union membership	-	-	-	3
Total number of items preceding module E:				**141**	**86**	-	**36**
Total number of items:				**269**	**214**	**130**	**147**

Notes: [1] In splitting module F into two parts, it was necessary to repeat a small number of questions and add some administrative items. For this reason, the sum of X and Y exceeds the total number of items in F.

Table A2 – Question wording of items analysed by set

Item Set	Question Wording	Response Categories
1) Agree/Disagree set 1		
TE4	"I'm always optimistic about my future."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE5	"In general I feel very positive about myself."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE6	"At times I feel as if I am a failure."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE7	"On the whole my life is close to how I would like it to be."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
2) Agree/Disagree set 2		
TE23	"I feel I am free to decide for myself how to live my life."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE24	"In my daily life, I seldom have time to do the things I really enjoy."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE25	"In my daily life I get very little chance to show how capable I am."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE26	"I love learning new things."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE27	"Most days I feel a sense of accomplishment from what I do."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree

		5=Disagree strongly
TE28	"I like planning and preparing for the future."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE29	"When things go wrong in my life, it generally takes me a long time to get back to normal."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE30	"My life involves a lot of physical activity."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
3) Agree/Disagree set 3		
TE40	"I generally feel that what I do in my life is valuable and worthwhile."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE41	"If I help someone I expect some help in return."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE42	"The way things are now, I find it hard to be hopeful about the future of the world."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE43	"There are people in my life who really care about me."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE44	"For most people in [country] life is getting worse rather than better."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
TE45	"I feel close to the people in my local area."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly

4) Agree/Disagree set 4		
TE53	"Considering all my efforts and achievements in my job, I feel I get paid appropriately."	1=Agree strongly, 2=Agree, 3=Neither agree nor disagree 4=Disagree 5=Disagree strongly
5) None or almost none/ all of the time set 1 (4-pt scales)		
TE8	How much of the time during the past week did you feel depressed?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE9	How much of the time did you feel that everything you did was an effort?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE10	How much of the time during the past week was your sleep restless?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE11	How much of the time did you feel happy?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE12	How much of the time during the past week did you feel lonely?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE13	How much of the time did you enjoy life?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE14	How much of the time did you feel sad?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?

TE15	How much of the time did you feel you could not get going?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE16	How much of the time did you have a lot of energy?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE17	How much of the time did you feel anxious?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE18	How much of the time did you feel tired?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE19	How much of the time were you absorbed in what you were doing?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE20	How much of the time did you feel calm and peaceful?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE21	How much of the time in the past week did you feel bored?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
TE22	How much of the time did you feel rested when you woke up in the morning?	1=None or almost none of the time 2=Some of the time 3=Most of the time 4=All or almost all of the time?
6) None of the time/ all of the time set 2 (7-pt scales)		

TE33	How much of the time spent with your immediate family is enjoyable?	0=None of the time 6=All of the time
TE34	How much of the time spent with your immediate family is stressful?	0=None of the time 6=All of the time
7) None of the time/ all of the time set 3 (7-pt scales)		
TE50	How much of the time do you find your job interesting?	0=None of the time 6=All of the time
TE51	And how much of the time do you find your job stressful?	0=None of the time 6=All of the time
8) Not at all/ a great deal set 1 (7-pt scales)		
TE35	To what extent do you get a chance to learn new things?	0=Not at all 6=A great deal
TE36	To what extent do you feel that people in your local area help one another?	0=Not at all 6=A great deal
TE37	To what extent do you feel that people treat you with respect?	0=Not at all 6=A great deal
TE38	To what extent do you feel that people treat you unfairly?	0=Not at all 6=A great deal
TE39	To what extent do you feel that you get the recognition you deserve for what you do?	0=Not at all 6=A great deal
9) Extremely satisfied/ dissatisfied set 1 (11-pt scales)		
TE31	How satisfied are you with how your life has turned out so far?	0=Extremely dissatisfied 10=Extremely satisfied
TE32	How satisfied are you with your present standard of living?	0=Extremely dissatisfied 10=Extremely satisfied
10) Extremely satisfied/ dissatisfied set 2 (11-pt scales)		
TE48	All things considered, how satisfied are you with your present job?	0=Extremely dissatisfied 10=Extremely satisfied
TE49	And how satisfied are you with the balance between the time you spend on your paid work and the time you spend on other aspects of your life?	0=Extremely dissatisfied 10=Extremely satisfied

Notes: [1]Question numbers shown are from the original ESS3 questionnaire.