![ISER - Institute for Social & Economic Research logo]

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

Annette Jäckle, Caroline Roberts and Peter Lynn

Acknowledgement:

Readers wishing to cite this document are asked to use the following form of words:

> Jäckle, Annette, Caroline Roberts and Peter Lynn (August 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.

The on-line version of this working paper can be found at http://www.iser.essex.ac.uk/pubs/workpaps/

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates

- MISOC (the ESRC Research Centre on Micro-social Change), an international centre for research into the lifecourse, and

- ULSC (the ESRC UK Longitudinal Studies Centre), a national resource centre to promote longitudinal surveys and longitudinal research.

ABSTRACT

This report presents findings from an experimental study carried out in the context of the European Social Survey, to assess the impact a change in data collection mode from the current face-to-face interviewing to telephone might have on data quality and to study the likely causes of any observed mode effects. Evidence from previous studies suggests which differences in response can be expected between telephone and face-to-face interviewing and also suggests likely causes of such differences. Previous empirical studies are, however, often limited in their ability to isolate different causes. The experimental design enabled us to distinguish mode effects caused by differences in the type of question stimulus used in each mode (audio vs. visual) and mode effects caused by the presence or absence of the interviewer. The design included three comparison groups (two interviewed face-to-face (one with showcards, one without) and the third by telephone). We found evidence of effects caused by the presence of the interviewer, but few stimulus effects. We tested a number of hypotheses about the likely causes of mode effects on response, focusing on three forms of satisficing and social desirability bias. We found no evidence that using showcards influenced response quality, either positively or negatively. Unlike previous studies, we found no support for the hypothesis that telephone respondents were more likely to satisfice. However, consistent with our expectations, we did find telephone respondents were more likely to give socially desirable responses across a range of indicators.

NON-TECHNICAL SUMMARY

The European Social Survey currently insists on face-to-face interviewing as its sole mode of data collection. However, owing to the mounting costs of carrying out face-to-face interviews and the divergent traditions and experiences of survey research across the different countries participating in the survey, there is a growing need to explore the alternatives offered by mixed mode data collection designs. Even relatively simple mixed mode designs – such as a switch to telephone interviewing in a small number of countries – could however threaten data quality, disrupting the continuity of the time-series for the countries concerned, as well as affecting the validity of cross-cultural comparisons.

The present experimental study is part of an ongoing methodological programme of research designed to inform decisions by the European Social Survey on whether to move to a mixed data collection strategy in the future, and if so, which modes to mix and how. The aim of phase II, the object of this report, was to assess the likely impact of a switch to telephone interviewing on data quality and to investigate the causes of mode effects in order to identify ways of mitigating these.

The main differences between face-to-face and telephone interviewing are the channels of communication and the physical presence of the interviewer. In a face-to-face setting, showcards can be used to make it easier for the respondent to understand questions and remember response categories. Over the telephone this aid is not available, making the response task more challenging. Similarly, the physical presence of the interviewer means that a range of non-verbal channels of communication are available. The interviewer may detect signs of waning motivation or misunderstanding and frustration on the part of the respondent and react to these more easily than over the telephone. Finally, face-to-face respondents are less likely to be engaged in other activities while answering survey questions and interviews are typically carried out at a slower pace than over the telephone.

As a result of these differences between the two modes, telephone respondents are likely to make less effort in answering survey questions (referred to as satisficing), resulting in different response distributions. For example, telephone respondents are more likely to say 'yes' or 'agree' and more likely to choose the same answer category for batteries of questions using the same scale. Face-to-face respondents may in turn report sensitive behaviours or attitudes less truthfully, since they will be more aware of the interviewer's reaction to their answers than a telephone respondent would be. As a result, face-to-face respondents may be more likely to edit responses to appear in a more favourable light (referred to as social

desirability bias). On the other hand, it may be easier for interviewers to establish rapport in a face-to-face setting. As a result the respondent might feel more comfortable reporting socially undesirable behaviours or attitudes.

Although previous studies have tested differences in responses across modes, their ability to infer the likely causes of differences was often limited. It is, for example, often not possible to distinguish whether the observed mode differences are a function of characteristics of the question (including question wording and response alternatives or the degree of sensitivity or complexity), characteristics of the mode (such as the presence or absence of an interviewer or the channel of communication (visual or aural) of the question stimulus and response) or characteristics of the respondent (such as propensity to satisfice or to give socially desirable responses).

Our study enabled us to distinguish mode effects caused by differences in the type of question stimulus used in each mode (audio vs. visual) and mode effects caused by the presence or absence of the interviewer. Since the European Social Survey relies heavily on the use of showcards, disentangling these effects is particularly important. The design included three comparison groups: two interviewed face-to-face (one with showcards, one without) and the third by telephone.

Mode significantly affected response distributions for over a third of the items tested. The differences between modes appeared to be small however, and did not affect the overall relationships between variables. Since the items included in the experiment were those deemed most sensitive to mode effects, the findings suggested that a switch to telephone mode might not affect the conclusions analysts would draw from the ESS data.

Most differences appeared to be due to the presence of the interviewer rather than the sensory channel, since mode effects were observed between face-to-face and telephone modes, but not between the two face-to-face groups. In general, we found no evidence that using showcards influenced response quality, either positively or negatively. This suggests that the ESS showcard questions were successfully adapted for the use over the telephone, by keeping modifications to a minimum. The main problems arose for the adaptation of numerical questions (about household income and hours spent watching television) which are formulated as banded questions in the ESS. Changing these to open-ended questions for the telephone resulted in large differences in response distributions.

Unlike previous studies, we found no support for the hypothesis that telephone respondents are more likely to satisfice. This suggests that the presence of the interviewer neither affected the difficulty of the response task nor the effort made by respondents. The

experimental survey was, however, much shorter and more varied than the full ESS survey and the possibility that telephone interviewing could lead to more satisficing can therefore not be excluded.

The most notable finding was that telephone respondents were more likely to give socially desirable responses across a range of indicators. This suggests that the advantages of trust built up in the face-to-face interview outweighed any disadvantages due to the lack of anonymity. In order to mitigate this effect, more research is, however, needed to understand the cognitive processes underlying social desirability bias. The traditional theory suggests that social desirability is the result of deliberate editing of responses. In our study, however, social desirability bias was more prevalent with telephone interviewing, although these interviews were conducted at a faster pace suggesting that respondents did not take additional time to edit their responses. Instead, respondents may have selected the most socially desirable response because it was the easiest, most accessible or salient response available to them without expending much effort on answering the survey question. Depending on the cause, the implications for reducing the impact of social desirability bias are clearly very different. Finally, little is known about cultural differences in social desirability bias and the extent to which our findings would replicate for other countries participating in the ESS. Differences may exist in the connotations of particular subjects, the social norms governing different types of behaviour and the importance of impression management strategies.

Table of Contents

1   Introduction

This paper provides a report on phase II of the ESS-Gallup Mixed Mode Methodology Project. This programme of methodological research into mixed mode data collection is being conducted by the Central Co-ordinating Team (CCT) of the European Social Survey (ESS), in collaboration with Gallup Europe and the University of Essex. Phase II of the research consisted of two mixed mode survey experiments conducted in Hungary and Portugal, designed to compare face-to-face interviewing (as is currently practiced on the ESS) with telephone interviewing (a preferred data collection option in many ESS participating countries). The aim of the experiment was to examine the likely impact on data quality of a switch to telephone interviewing, focusing in particular on separating two different types of mode effect: those due to differences in the sensory channels used to communicate information and those due to the presence of the interviewer. In this report, we describe the background to the programme of research, the rationale behind the design of the phase II study and the theoretical impetus for the analysis we have undertaken. We present the results of the analysis of data collected in Hungary[1] and provide preliminary recommendations based on the findings for future data collection methodology on the ESS, as well as the future direction of this programme of research.

2   Background

2.1   Data collection on the ESS – a mixed mode future?

Cross-national surveys are faced with a number of challenges not typically encountered in national studies; to ensure comparability of the datasets across different countries, they depend for their reliability on a sort of 'principle of equivalence'(Jowell 1998), which applies to all aspects of the survey process – sampling (Lynn et al 2004), question wording, response options, coding schema and so on. For this reason the most ambitious multinational projects tend to require all participating countries to employ the same mode of data collection. In the case of the ESS – a multi-nation project designed to measure and monitor changing social values in Europe – the exclusive mode for data collection is face-to-face interviewing.

The decision to opt for face-to-face interviews over alternative modes of data collection was driven by a number of different factors. Face-to-face interviewing has long been recognised as a kind of 'gold standard' among data collection methodologies. It has

---

[1] Delays to the completion of fieldwork in Portugal meant that the data were not available in time to include in the present analysis.

demonstrated advantages with respect to obtaining high response rates (de Leeuw 1992; Holbrook, Green, and Krosnick 2003), because of the effectiveness of in-person contacts at persuading would-be respondents to take part. It is also credited with obtaining better quality data compared with telephone interviews or self-administered modes, partly because the interviewer (particularly when using computer-assisted personal interviewing, or CAPI) is able to ensure that questions are not accidentally skipped, that respondents' answers are fully probed if necessary and that answers are recorded correctly (particularly with long and complex questionnaires such as those used on the ESS). Moreover, in the comparative European context, in-person interviewing offers several other important advantages, making it the best-possible *unimode* data collection option for a rigorous cross-national survey. In particular, cross-cultural variations in literacy levels would have prohibited the sole use of self-administered questionnaires, while national differences in the extent of population coverage offered by alternative data collection technologies (such as fixed-line telephones) precluded their use.

Yet despite its advantages, the sole use of face-to-face interviewing may not in fact be the best data collection strategy for the ESS in the longer term. The cost of conducting fieldwork using face-to-face interviews varies widely across the 31 countries that have so far participated in the survey. For some geographically larger countries such as Sweden or Norway, for example, it is increasingly regarded as an unaffordable luxury. Other modes are much cheaper than face-to-face and by no means proportionately inferior in quality. Entrenched variation in fieldwork practice and preferences across countries act as a further barrier to insisting on a single data collection strategy. Telephone interviewing – which long took over from face-to-face interviewing in the United States as the dominant survey data collection mode – is similarly popular, and practiced to a high standard in a number of European countries (so much so that in some cases, it is increasingly becoming the only *viable* option, because the choice of survey organisations prepared and equipped to undertake in-person interviews has become so limited). Such variations in survey practice are further reflected in public preferences for different data collection methods in different countries, which, in turn, may have consequences for levels of participation. For several countries taking part in early rounds of the ESS, response rates fell well below the target of 70% and the costs of trying to maintain them using face-to-face methods alone may mean that some countries will find it increasingly difficult to continue participating in the survey. As with other aspects of cross-cultural survey methodology (such as sampling and translation),

insisting on the *same* methods, may not be the best way to ensure that *equivalent* data are obtained.

A variety of mixed-mode data collection designs might offer a solution, however. One possibility, for example would be to specifically incorporate variations in data collection mode at different stages of fieldwork; for example, in a 'sequential' design where non-respondents to the primary (usually most cost-efficient) mode are re-contacted using an alternative (more costly) mode and so on (Hochstim 1967; Japec 1995). Such a strategy has been shown not only to reduce costs (Dillman 2000; Groves 1989), but also the possibility of non-response bias (Mooney, Giesbrecht, and Shettle 1993; Moore 1998; Schaefer and Dillman 1998). An alternative method of incentivising participation (though still a somewhat distant option in comparative surveys) would be to offer all sampled respondents in each country a choice of (say) four available modes of data collection from which they could select their preferred option – face-to-face, telephone, web, or paper self-completion (Dillman et al. 1995; Dillman, Clark, and West 1995; Dillman et al. 2000; Groves and Kahn 1979). The empirical findings in support of such an approach have so far been less compelling, however (de Leeuw 2005). A more realistic option for the moment is to allow survey agencies to choose their own preferred method as the *most appropriate* data collection mode for that country (in terms of achieving high quality data at affordable cost) – which would mean that the most likely mixed-mode scenario (if any) for the ESS in the near future, would be for countries meeting appropriately stringent quality criteria to switch to telephone interviewing for the survey, either instead of, or in conjunction with, face-to-face interviews.

At present, we cannot be sure whether – or how best – viable alternatives to face-to-face interviewing might be introduced successfully alongside face-to-face interviewing in future rounds of the ESS, without damaging equivalence across countries and over time.

2.2   Challenges involved in mixing modes

Differences in the data that come about because of the chosen mode of data collection are referred to as mode effects. They can take the form of 1) coverage error, 2) non-response bias or 3) measurement error.

The first error component refers to differences in the definition of the population of inference due to mode of data collection. The ESS aims to cover the residential population aged 15+ in each country. Coverage error results if some members of this population have a zero chance of being included in the sample. The extent to which different modes are

susceptible to coverage error differs across countries (Lyberg and Kasprzyk 1991). For example, telephone penetration varies across the countries participating in the ESS and may be near universal in some countries but very low, say, among the rural populations in others. In addition, the availability of complete lists of telephone numbers which can be used as sampling frames varies. As a result of such practical constraints, the characteristics of the population covered might differ across countries, hampering comparative research (Braun 2003). On the other hand, a mix of modes for which complementary sampling frames exist could be used within countries to improve population coverage.

The second component is non-response bias associated with data collection mode. To the extent that people have different mode preferences (response rates are, for example, consistently lower with telephone than face-to-face interviewing), the use of different modes across countries could lead to differential non-response. That is, non-respondents may not only differ from respondents in characteristics of interest to the study, but this difference may be different across countries, again hampering comparative research. On the other hand, a mix of modes within a country may make it possible to access a greater variety of people and thereby reduce the problem of bias associated with non-response.

The third component of mode-related error (which Saris (1997) refers to as 'transformation error') is that of measurement error. This is the more familiar manifestation of mode effects and refers to the finding that the mode can have an impact on the data collected, sometimes generating different response distributions (de Leeuw 1992; Dillman et al. 1996; Hochstim 1967; Schwarz, Hippler, and Noelle-Neumann 1992). Differences between modes in measurement error tend to vary (in nature and/or magnitude) according to the types of question being asked (Aquilino 1994; de Leeuw and van der Zouwen 1988; DeMaio 1984; Dillman and Tarnai 1991), the order in which they are asked (Bishop et al. 1988) and the types of response options offered (Krosnick and Alwin 1987; Schuman and Presser 1981).

Modes are likely to lead to differences in responses, if they have differential effects on the ways in which respondents come up with an answer. The quality of a response is determined by how carefully the respondent executes the process of understanding the question, retrieving information (including feelings, beliefs and knowledge about an issue), integrating information to form an overall judgement and formulating a response (Tourangeau, Rips, and Rasinski 2000). Whether the respondent executes this process well and truthfully depends on whether he or she makes 'sufficient' effort and is willing to disclose sensitive or potentially embarrassing information. Which level of effort is sufficient

depends on a combination of the task difficulty and the respondent's ability and motivation. There is evidence, for example, that respondents are more likely to shortcut the first three stages of the response process if the task is difficult or respondents have low cognitive ability or motivation (see Krosnick 1991; 1999). Similarly, the willingness to self-disclose and to report truthfully is likely to depend on a combination of the perceived sensitivity and resulting threat of the question, the legitimacy of the survey and the degree of privacy or anonymity of the reporting situation. Respondents are more likely to edit their responses and not report truthfully, if they are uncomfortable about the confidentiality of sensitive information or about social sanctions to undesirable answers.

Deviations from the optimal response process are referred to as 'satisficing' in the case of shortcutting (Krosnick 1991) and as 'social desirability bias' in the case of editing responses to appear in a more favourable light (see DeMaio 1984). Satisficing may be of a weak form, in which case respondents merely execute the different stages of the response process less completely, or may be of a strong form, in which case respondents may skip some of the stages altogether.

If modes have differential effects on the factors determining whether the respondent's effort is 'sufficient' or on the willingness to self-disclose, then modes are likely to lead to differences in the execution of the response process, and hence to differences in the quality of responses. For example, if the task of answering a particular item is more difficult in one mode than in another, then it is more likely that respondents in the first mode will shortcut the response process. Krosnick (1991) argues that the different factors interact and that, for example, increased task difficulty is likely to lead to more shortcutting among low ability or low motivation respondents. If mode affects task difficulty and respondent motivation, then the effects on the response process are likely to be magnified for low ability respondents.

In sum, if the ESS were to consider allowing (or encouraging) mixed mode data collection, a specific concern would be that mode-related errors might be differential across countries, thus confounding comparisons. Similarly, switches between modes over time could confound the measurement of change in the time series. In reality it may, however, be necessary to accept the trade off between different sources of survey error. If face-to-face interviewing remained the single mode of data collection, the ESS would risk obtaining response rates as low as say Switzerland's in round 1 of the ESS (32%). The potential damage to the comparability across countries or the continuity of the time-series caused by mixing modes might not appear so bad in comparison. In order to evaluate this trade off, the

ESS requires information about the magnitude and nature of the different sources of error in the survey data and the effects a change in modes would have.

3 ESS-Gallup mixed mode methodology project

A modest budget for methodological work was built into the contract for the first two rounds of the ESS and the decision was taken to invest this budget in a programme of research investigating the feasibility of changing the current ESS policy of single-mode data collection using face-to-face interviews, to a mixed mode data collection strategy in its future rounds.

The aim of this ongoing programme is to provide information that will help to inform decisions by the ESS regarding:

- whether mixed-mode data collection should be allowed on future rounds of the ESS;
- which modes should be allowed;
- within which kinds of overall survey design mixed modes could be employed.

This information will include assessment of the following issues in the context of a survey instrument similar to the current ESS interview questionnaire:

- coverage and response rates that can likely be achieved with different modes and mode combinations;
- likely differential error between modes (particularly non-response error and measurement error) and its causes.

To date, the research has focused on gathering information about the latter. The research we have undertaken has been carried out in conjunction with the Gallup Organisation, Europe. This collaboration has allowed the Central Co-ordinating Team of the ESS to benefit from a larger-scale research project than would have been possible alone, as it has entailed the pooling of financial resources (with both parties contributing 50%). The project has consisted of two phases:

3.1 Phase I

The first phase of data collection took place in May and June 2003 in Hungary. The study consisted of a 'hall test', in which participants, selected by quota sample to be representative of the Hungarian urban population by age, gender and education, were randomly assigned to

one of four interview conditions: face-to-face interview, telephone interview, self-completion paper and pencil questionnaire and web-based questionnaire. Participants were then re-interviewed in a different mode[2]. All participants received the same stimuli questions in each of the four interviewing modes, making it possible to compare responses to different types of survey question across pairs of modes, and to examine differences in responses both between and within participants.

The analysis of the phase I experiment (Peytcheva et al. 2004) indicated a number of areas that we felt merited closer attention and the design for the second phase was drawn up in light of its conclusions. In particular, the following were important factors influencing the focus of phase II of the research:

1. Responses to the telephone interviews in the phase I experiment were found to differ greatly from responses in each of the other modes. Given that a switch to telephone interviewing – either in addition to or instead of face-to-face interviewing in certain countries - represents the most likely scenario for mixing modes on the ESS in the foreseeable future, it was felt that the second phase of research should explore further the particular problems associated with conducting the ESS interview by telephone and to focus on comparisons between data collected by telephone and face-to-face.

2. Consistent with the findings of other mode studies, certain types of questions were found to be more susceptible to mode effects than others. These included attitudinal items using scales and sensitive items that were subject to social desirability bias in the two interviewer-administered modes. There was also evidence of response order effects consistent with whether the question stimulus was visual or aural (including where showcards were used in the face-to-face interview). Nevertheless, the experimental design was such that question characteristics were sometimes confounded with each other, making it difficult to discern the type and likely cause of the observed mode effects and as a consequence, to recommend possible ways of mitigating effects. It was concluded that the design of the second phase of the research should enable the analysis of the types of question most susceptible to differential measurement error across modes and the likely causes of mode

---

[2] The experimental design was not fully balanced, however, and those interviewed by web in the first wave of data collection were not re-interviewed.

differences, so as to inform recommendations about mitigating effects in a mixed mode ESS.

3. Of all the items in the questionnaire, statistically significant differences between the face-to-face and telephone conditions were more frequently observed on those that needed to be redesigned for administration by telephone, because they relied on complex showcards in the face-to-face mode. There was also evidence that modifications to response categories and question formats for other questions in the experiment influenced the magnitude of the observed mode effects. It was decided, therefore, that the next stage of experimentation should focus on the particular problems of adapting the ESS face-to-face questionnaire for use in telephone interviews.

4. A number of features of the experimental design made it difficult to draw robust conclusions about the precise cause of differences in response distributions by mode. In particular, it was not always possible to determine whether the mode effect was a function of characteristics of the question (including question wording and response alternatives, as well as degree of sensitivity or complexity), characteristics of the mode (such as the presence or absence of an interviewer or the channel of communication (visual vs. aural) required by the question stimulus and the response) or characteristics of the respondent (e.g. propensity to satisfice or to give socially desirable responses, etc.). Thus, there were strong arguments for narrowing the focus of the research at phase 2 and in so-doing, reducing the complexity of the experimental design.

5. While the design carefully avoided confounding differential coverage or non-response error with differential measurement error (as the random allocation took place *after* a subject had agreed to take part), this came at the price of sacrificing some realism. For example, telephone respondents were not in their own home with the possibility of other household members being present; instead, they were in a booth in a hall and had already met a researcher face-to-face.

The intention at the start of the research programme was to carry out at least one further phase of data collection, involving between three and five other countries. In the end, due to

budgetary restrictions it was only possible to conduct one further phase of experimental work. However, it was agreed that the research should address the question of whether there are cross-cultural differences in the differential measurement errors observed between modes. The phase II study was therefore carried out in two ESS participating countries.

3.2   Phase II

The objective of phase II was to address particular problems associated with administering the ESS interview by telephone. It involved a direct comparison between the current face-to-face methods employed on the ESS and telephone alternatives. The principal aims were:

1. To assess the likely *impact* of a change to telephone interviewing on data quality in the ESS: which questions on the core questionnaire would be most likely to be sensitive to mode of interviewing and what the nature and magnitude of mode effects would be.
2. To identify *causes* of observed mode differences, in particular to disentangle effects arising from differences in the question stimuli (i.e. how questions are asked in different modes) and effects arising from the physical presence of the interviewer (and other characteristics of the mode not specifically controlled for).
3. To make recommendations about how to *mitigate* mode effects, for example through modifications to question wording and response formats, interviewer instructions, etc.

The research was also designed to begin to examine some of the issues concerned with conducting interviews with respondents on their mobile (cell) phones and to explore cross-cultural differences in response errors. In the present report we focus on the comparison of face-to-face interviewing with (fixed) telephone only.

4   Method

4.1   Research design

Two experiments with the same design were conducted by Gallup Europe in Hungary and Portugal starting in July 2005. The experiments consisted of interviews conducted face-to-face in respondents' homes and telephone interviews conducted by fixed-line telephone (also in respondents' homes) or by mobile phone[3]. The interviews consisted of a subset of questions from the core questionnaire of the European Social Survey. In order to reduce costs, the fieldwork was concentrated in the countries' capital cities (Budapest and Lisbon), which also offered the advantage of suitable sampling frames in both locations, including telephone numbers *and* addresses, thereby holding error from sampling/ coverage constant between modes. Fieldwork was carried out by the Gallup Organisation Hungary in Budapest and by its partner organisation, Consulmark in Lisbon.

Our primary interest was in comparing data from face-to-face survey interviews such as those conducted on the ESS with data from telephone interviews. However, most mode comparison studies that have looked at face-to-face and telephone interviews have neglected to control for how questions are asked in each mode (Holbrook, Green, and Krosnick 2003), making it difficult to draw robust conclusions about the presence of mode differences. Confounding the influence of the question stimulus with other features of the mode also restricts the conclusions that can be drawn about the likely causes of any observed effects. Our experiment was designed to allow us to isolate the primary causes of inter-mode differences with a view to developing recommendations for how to mitigate their effects in a mixed mode ESS. In particular, we wished to distinguish between what we refer to as 1) 'stimulus effects', resulting from differences in the question form or medium in which the response categories are communicated (e.g. whether or not showcards are used); and 2) mode effects *per se*, resulting from other features of the mode – notably, the presence or absence of the interviewer, but also other aspects such as the pace with which the interview is conducted and the impersonality, legitimacy and cognitive burden imposed by each mode (Tourangeau, Rips, and Rasinski 2000) . Thus, in order to ensure strict comparability between the modes (and thus isolate mode effects *per se* from stimulus or question wording effects), we included a third treatment condition in the experiment, which used the same questionnaire in a face-to-face interview as that used over the telephone (i.e. one without showcards).

---

[3] The analysis reported here is based on data from telephone interviews conducted by fixed line only.

The experimental design therefore included three treatment groups. Two groups were interviewed face-to-face and the other was interviewed in telephone mode[4] as follows:

Group 1   Face-to-face interview with showcards

Group 2   Face-to-face interview without showcards

Group 3   Telephone interview (fixed-line and mobile)

4.2 Questionnaires

In order to ensure continuity of the time series, only minimal changes to the current ESS face-to-face methodology would be permissible in the event of a switch in data collection mode in certain countries. We therefore wished to compare the alternative mode (telephone interviewing) with the typical ESS interview method. For this reason, the experiment required two versions of the questionnaire. For group 1, the question and showcard design was essentially identical to the ESS[5] round 2 questionnaires (though the questionnaire was much shorter). For groups 2 and 3, the questions using showcards were modified so that they could be administered orally, either by telephone or in a face-to-face interview without showcards. A number of different adaptations were made: 1) for most questions, the interviewer either provided a description of the response scale to be used or read out the available response categories; 2) for a limited number of more complex items or questions with long lists of response alternatives, the format was changed to make them more suitable for use in a telephone interview. This involved either a) breaking the original question down into sub-questions (e.g. for items classifying occupation); b) converting the question into an open-ended format, in the case of items asking about quantity (e.g. of time; income); or c) reducing the number of response categories (this affected one item only). Responses to all items in the questionnaire could be coded by the interviewer as 'refused' or 'don't know', as is standard practice on the ESS; 'no opinion' was not offered *explicitly* to respondents.

Items in the questionnaire were selected from the ESS to provide a variety of indicators of data quality across each of the two modes, based on those used in other mode comparison studies (de Leeuw 1992; de Leeuw and van der Zouwen 1988). A further

---

[4] The telephone group consisted of two subgroups, split between fixed-line and mobile phones, to allow us to compare differences between the modes of telephone interviewing. These comparisons are not addressed here.

[5] Except where translation errors were found in the Hungarian ESS questionnaires, in which case it was necessary to modify the translation because of the requirement for equivalence with a comparable experiment conducted in Portugal as part of the same programme of research.

criterion was to select items that were believed to be most likely to be susceptible to mode effects, so that we can draw conclusions about the overall mode sensitivity of the ESS questionnaire. Most of the items selected were measures of social attitudes and values. We included questions on social and political trust, political self-efficacy, life satisfaction, trust in institutions, religiosity, attitudes towards immigration, gender roles, gay and lesbian rights and obedience to the law. We also included the following behavioural measures: time spent watching television daily; time spent watching news programmes; voting and party voted for; and frequency of religious service attendance (see appendix A for question wording). In addition, the questionnaire included a range of socio-demographic measures (including sex, age, education, occupation and income).

Our assessment of the susceptibility of items to mode effects was based either on the nature of the question topic or the format of the response scales or categories. We selected measures that have been shown elsewhere to be sensitive to social desirability bias or to have widely shared social desirability connotations, such as income (de Leeuw and van der Zouwen 1988), religious service attendance (Groves 1989), voting (Hadaway, Marler, and Chaves 1993; Karp and Brockington 2005; Krosnick 1999; Silver, Anderson, and Abramson 1986) time spent watching television news and interest in politics (Cassel 2004). We also included questions on topics about which we thought respondents might feel unwilling to disclose their true opinions in order to portray themselves to interviewers in a favourable light (such as immigration and equality between the sexes). Of these, we believed we had strong grounds for including the items on immigration, as a number of US studies have demonstrated the social desirability connotations of questions measuring racial attitudes, particularly among white respondents (e.g. Groves and Kahn 1979; Holbrook, Green, and Krosnick 2003). However, no attempt has been made to establish empirically which items on the ESS might be most susceptible to social desirability bias, so our research shares this weakness with other studies that have assessed mode differences in socially desirable reporting (Krysan et al. 1994).

We also selected questions that would act as indicators of respondent satisficing. The questions selected used a variety of response formats; we included questions with nominal response categories that we thought might be susceptible to response order effects and batteries of scale items that might be susceptible to non-differentiation and (for those with agree/disagree scales) acquiescence. We also included a number of items designed to measure respondents' experience of the interview such as perceptions of interview length, mode preference and whether the respondent had felt uneasy answering questions on the

topics covered in the questionnaire (based on the method used by Groves and Kahn (1979)), the latter intended as a crude measure of the social desirability connotations of some of the questions.

In order to explore some of the issues associated with mobile phone interviews, respondents in the telephone mode were asked 3 additional questions - about where they were at the time of interview and what they were doing (extent of multi-tasking), and about their telephone type (whether mobile or fixed line, and if the latter, whether it was portable or not). Finally, interviewers were asked to complete 6 questions about the respondent (whether they seemed distracted, whether they made an effort to answer questions, whether they had difficulty understanding questions, whether others were present during the interview, and so on). The purpose of these questions was to provide further background information that might be associated with differences between modes.

All questionnaires were devised in (British) English and were translated into Portuguese and Hungarian. However, questions in the group 1 questionnaire were extracted from the ESS source questionnaire and translations were already available for all items from rounds 1 and 2 of the ESS.

4.3 Sampling and response rates

As stated, fieldwork took place in the Greater Budapest region, which offered the advantage of a single sampling frame, including telephone numbers *and* addresses, thereby holding any error from sampling/ coverage consistent across the experimental groups. An equal-probability sample of fixed residential phone numbers within the defined area was selected. Each unit was randomly allocated to one of the 3 treatment groups. At each contacted household, one person aged 15 or over was randomly selected for interview using the last birthday method.

The *target* responding sample sizes were 500 in each of the two face-to-face treatments and 1000 in the telephone group, of which 500 minimum were to be interviewed on a fixed line. Selected sample sizes were calculated according to expected response rates in each of the groups based on other face-to-face and telephone surveys.

For all sampled units allocated to the face-to-face groups, first contact was to be by telephone. However, the field agency found this method of arranging in-person interviews to be unsatisfactory, so it was necessary to change the mode of contact during fieldwork to ensure adequate levels of participation. In total, 3300 units were contacted in the face-to-face

groups at the first stage (of which 1852 were *initially* contacted by telephone and 1448 were contacted in person by the interviewer). In total, 515 respondents were interviewed face-to-face using showcards, and 518 respondents were interviewed face-to-face without showcards. Table 1 provides a full break-down of the issued sample.

Table 1: Interview outcomes

	N
Total numbers sampled	6150
Total allocated to be interviewed face-to-face	3300
Total allocated to be interviewed by telephone	2850
Telephone: Total numbers dialled	**2850**
No contact	147
Ineligible	102
Refusals	858
Unknown eligibility	530
Other eligible, non-interview	146
Completed screener	1067
Assigned to fixed-line interview	*685*
Assigned to mobile interview	*382*
Broke off prior to interview	*180*
Face-to-face: Total sampled units	**3300**
Telephone contact only attempted	627
Mixed telephone and face-to-face contact attempted	1225
Face-to-face contact only attempted	1448
No contact	786
Ineligible	125
Refusals	1204
Unknown eligibility	35
Other eligible, non-interview	117
Completed interviews	**1920**
Face-to-face with showcards	515
Face-to-face without showcards	518
Fixed-line telephone	685
Owns mobile phone	*342*
No mobile phone	*343*
Mobile telephone	202

For units allocated to the telephone group, the interviewers attempted a short screening interview with the selected respondent to establish whether he or she had a mobile telephone. Half the respondents with mobile phones who agreed to participate were randomly assigned to be interviewed on their fixed-line and the remainder were asked to give their mobile telephone number, so they could be re-contacted to complete the interview on their mobile. Respondents assigned to the fixed-line group who agreed to participate were immediately administered the telephone interview. In total, 1067 (out of an initial sample of 2850)

14

respondents completed the screening questions and 887 completed the telephone interview. Of these, 685 were interviewed on their fixed-line telephone. In analysis, owners of mobile phones in the fixed-line group were weighted by two to adjust for the fact that only half of this sample was interviewed on a fixed-line telephone. The response rate for the telephone group was 32% and that for the face-to-face groups combined was 33% (AAPOR 2006response rate 1).

5 Analytical framework and related evidence from previous studies

We first examined the magnitude of mode effects and then tested specific hypotheses about the nature of differences and their causes. To shed further light on the differences caused by modes, we also assessed the respondents' experience of the interview in each of the different modes. This section describes the methods used to examine the magnitude of mode effects and the hypotheses, including how they were operationalised and the statistical methods used to test them. The findings are then presented and discussed in Section 6.

5.1 Magnitude of mode effects

We tested the magnitude of mode effects using a regression approach, in order to control for differences in the sample composition of the three treatment groups, in terms of their socio-demographic make-up (see section 6.1). Using OLS we compared responses at the mean and tested whether modes affected relationships between variables and summed attitude scores. OLS might however provide misleading results, since most items are ordinal and therefore the intervals between adjacent response categories cannot be assumed to be equal. We therefore also tested ordinal items using models where higher values of the dependent variable are assumed to correspond to higher outcomes, but the actual values are irrelevant. To do this, we used proportional odds modelling.

The proportional odds modelling technique (also referred to as cumulative odds model (O'Connell 2006) parallel regression model, or grouped continuous model (Long 1997) is, according to Billiet and Welkenhuysen-Gybels (2004) currently the best method available for assessing the measurement equivalence of attitudinal indicators. The authors use this technique to test for equivalence of immigration items (three of which were also carried in the present experiment) across 21 countries in the first wave of the ESS. The proportional odds model is equivalent to a sequential series of binary logistic regressions of $P(Y>j)$ over

cumulative splits of the data, where the coefficients are constrained to be equal in each equation:

$$P(Y_i > j) = \frac{\exp(\alpha_j + X_i\beta)}{1 + [\exp(\alpha_j + X_i\beta)]} \quad j = 1, 2, ..., M - 1$$

where *M* is the number of response categories of the ordinal dependent variable *Y*. For a variable with 4 response categories there are 3 cumulative dichotomisations *j* = 1, 2, 3: 1 versus 2, 3, 4, then 1, 2 versus 3, 4, then 1, 2, 3 versus 4, where the categories below and including *j* are recoded to '0' and the categories above j are recoded to '1' (see O'Connell (2006) for ordinal models representing alternative dichotomisations. Compared to separate estimations of each cumulative dichotomisation, the results differ slightly when all equations are estimated simultaneously (Williams 2006)).

The proportional odds model assumes that the odds(*j*) = P(*Y*≤*j*) / P(*Y*>*j*) have the same ratio for all combinations of explanatory variables for any dichotomisation *j*. This smoothness assumption implies that explanatory variables have the same effect on the odds, in each of the cumulative splits of the data: if telephone interviewing led, for example, to higher scores on an immigration item, this difference should reveal itself in each response category. This assumption may however not hold. Indeed the hypotheses about the causes of mode effects discussed below, predict that some response categories are likely to be affected differentially by mode.

An approach which does not obscure differential patterns of explanatory variables, but is more parsimonious than estimating the full set of cumulative logistic models, is to use partial proportional odds models. In this case explanatory variables for which the proportional odds assumption holds are constrained to be equal, while variables for which the assumption is violated are allowed to vary across the cumulative dichotomisations. This model is implemented in Stata by Williams (2006). The estimation uses a backwards stepwise selection procedure, starting with a model corresponding to the full set of cumulative logistic models and gradually imposing constraints for variables for which the assumption of proportional odds holds, based on Wald tests of the equivalence of coefficients across equations. We used this procedure and tested the proportional odds assumption for mode and all socio-demographic variables, using a (conservative) .01 level of significance for the decision to impose constraints.

For items where the proportional odds assumption holds, the standard error of the single mode coefficient can be used to test the significance of mode effects on the response

distribution. For all other items we test for the overall effect of mode using Wald tests of the joint hypothesis that all mode coefficients from the series of cumulative logistic dichotomisations equal zero, as well as reporting the significance of mode coefficients from each model.

Since the coefficients do not correspond to the marginal effects of changes in mode on the probabilities, their interpretation is somewhat limited. We therefore present the predicted probabilities for each response category and differences in distributions due to mode, where

$$P(Y_i = 1) = 1 - P(Y_i > 1)$$
$$P(Y_i = j) = P(Y_i > j_{-1}) - P(Y_i > j) \quad j = 2, ..., M - 1$$
$$P(Y_i = M) = P(Y_i > j_{M-1})$$

This provides a more complete picture of the effects of mode on response distributions and of divergent patterns which would otherwise be obscured or distorted by the proportional odds model.

5.2 Hypotheses about causes of mode effects

To assess the nature of mode effects, we test specific hypotheses about how the differences between modes may lead to differences in responses. Telephone and face-to-face interviews mainly differ in two respects: the sensory channel available for the transmission of information, in particular the possibility of using showcards in face-to-face interviews, and the physical presence of the interviewer. (For discussions of factors that may impact on responses in mode comparisons, see for example de Leeuw (2005), Holbrook, Green and Krosnick (2003), Fricker, Galesic, Tourangeau and Yan (2005), and Tourangeau and Smith (1996).) There are a number of potential explanations of how these differences between the two modes may affect task difficulty, respondent motivation and willingness to self-disclose and hence the response process and resulting responses. (See Krosnick 1991 for an extensive discussion of sources of task difficulty and respondent motivation, some of which may not be affected by mode.) Each of these potential explanations has empirical implications which we test in the following. The hypotheses we derive are mainly based on discussions by Krosnick, Narayan, and Smith (1996), Tourangeau, Rips, and Rasinski (2000), Holbrook, Green, and Krosnick (2003) and de Leeuw (1992; 2005) and are summarised in Table 2.

Table 2: Summary of hypotheses

	H1	H2		H3/H4	H5	H6	H7
Effect	Satisficing	Primacy	Recency	Satisficing	Satisficing	Social desirability	Social desirability
Comparison groups and expected direction	F2f sc < F2f nosc	F2f sc > F2f nosc	F2f sc < F2f nosc	Tel > F2f nosc	Low ability > High ability	Tel < F2f nosc	Tel > F2f nosc

Notes: F2f = face-to-face, Tel = telephone, sc = showcard, nosc = no showcard.

H1: Showcards simplify the response task, because the visual presentation reduces the burden on the respondent to remember response categories and may make it easier to understand the question. If this is true, then we should see less shortcutting with showcards than without. (Of course this assumes that respondent motivation remains constant: if we do not observe less shortcutting with showcards, it may be the case that in addition to simplifying the task, the showcards have the additional effect of reducing the respondents' motivation leading to a negative net effect.)

H2: The responses produced by shortcutting are likely to be different with aural and visual presentations. With showcards respondents are likely to read down the list until they find a plausible answer. With aural presentation respondents are more likely to remember the last response categories (this assumes that interviewers always read the complete list and are not interrupted by the respondent picking an earlier response). If there is shortcutting, then we should expect to see more responses from earlier categories ('primacy effect') with showcards and more responses from later categories ('recency effect') without showcards. We expect this effect to be stronger where long lists of response categories are used.

H3: The interviewer's presence reduces the task difficulty, because in a face-to-face situation the interviewer is able to make use of a range of communication channels in order to facilitate the respondents' comprehension of the survey task, and thereby reduce the cognitive burden. For example, the interviewer is able to respond to any signs of confusion of miscomprehension on behalf of the respondent by clarifying what the question is asking or how the respondent should answer. In addition, face to face interviews are typically administered at a slower pace than telephone interviews, possibly because telephone interviewers minimise awkward silences and face a higher risk of break-offs. As a result of the faster pace it may be more difficult for respondents to understand questions and they are likely to take less time for cognitive processing. Face to face respondents are also less likely

to be distracted and to be doing other things while answering the survey questions. Consequently, we expect to see more shortcutting with telephone interviewing than with face to face interviewing (without showcards). (But see also H4 below.)

H4: The interviewer's presence increases respondent motivation, because the respondent can observe nonverbal cues of the interviewer's commitment and enthusiasm, while the interviewer can detect nonverbal cues of declining motivation and react to these. As with H3, this would lead us to expect more shortcutting with telephone interviewing than with face to face interviewing (without showcards). Of course, if this is observed, we will not be able to distinguish between the interviewer's effects on task difficulty (H3) and on motivation (H4).

H5: The impact of mode is likely to be largest among respondents with low cognitive ability, because increased task difficulty and lower motivation are likely to be more detrimental for respondents with lower ability. If this is true, then we should see larger differences in the extent of shortcutting across modes for low ability respondents than for high ability respondents.

H6: The interviewer's presence reduces anonymity and 'social distance' and may make the reporting of sensitive information more threatening. The reason is that the respondent can observe nonverbal signs of approval or disapproval. The fear of such sanctions is likely to reduce the respondent's willingness to disclose sensitive or potentially embarrassing information. If this is true then we should see more socially desirable answers with face to face interviewing.

H7: The interviewer's presence improves the rapport with the respondent, because nonverbal communication aids the development of interpersonal trust. In addition, the possibility of showing the interviewer's identification materials can help establish the legitimacy of the survey questions. In comparison, telephone respondents may feel less confident that the interviewer will protect the confidentiality of their responses and may therefore be less willing to disclose personal information. If this is true then we should see more socially desirable answers with telephone interviewing. It is, however, not clear what the net effect of H6 and H7 is likely to be.

To test these potential explanations for mode differences, we compare the extent of shortcutting or respondent satisficing and social desirability responses across modes, both at the aggregate and individual item level. Note that a *response* that might be indicative of satisficing could also be the result of careful reflection. In the absence of mode effects, one would however expect the response *distributions* of two samples of the same population to be comparable (assuming comparability of other features of the surveys). In this context, if the extent of satisficing and social desirability responses varies across modes, this would supply evidence consistent with the above hypotheses. If instead the extent of satisficing and social desirability responses is unrelated to mode or varies in unexpected directions, the hypotheses that mode differences are due to satisficing and social desirability bias should be rejected.

5.3 Evidence from previous studies

The general conclusions from previous studies seem to be that differences between telephone and face-to-face surveys are small (see for example de Leeuw 1992; Kalton, Kasprzyk, and McMillen 1989). Groves (1989) concluded that for "estimates of population means and proportions, similar conclusions will be drawn from telephone and face to face interview surveys" (p. 514). Groves and Kahn (1979) estimated over 200 means on the total sample and found only a small number of differences between modes; Sykes and Hoinville (1985) estimated 95 means on the total sample and found differences for only 9; Körmendi (1988) examined 291 means and found only 11 significant differences. Holbrook et al. (2003), however, argue that although mode effects may appear small, they can be sizeable among the less educated and as a result can affect comparisons across subgroups.

Previous studies have tested combinations of the hypotheses about causes of mode differences discussed above, using similar indicators of satisficing and willingness to self-disclose. The following presents a brief summary of findings which provided evidence in favour or against our hypotheses.

H1: Relatively few studies have explicitly tested the benefits of using showcards for response quality, although a number of studies have considered the difficulties of administering by telephone questions for which showcards are used in face-to-face interviews (e.g. Miller 1984; Sykes and Collins 1988). It appears to be widely acknowledged that showcards can serve to facilitate the response process, by acting as prompts, aids to recall and even as a way of enhancing the privacy of the interview, such as where showcards display labelled response

options, and respondents are only required to read out the label corresponding to their answer (Duffy 2003; Miller 1984). There is also evidence that interviewers find showcards useful, possibly because they speed up the response process (Rogers 1976), but that respondents may feel pressure as a result to read response cards quickly. In this way, showcards may actually increase the level of cognitive burden on respondents (Duffy 2003; Sykes and Collins 1988) and serve to distract them (Dijkstra and Ongena 2002). There have been relatively few empirical tests of these alternative hypotheses looking specifically at how the use of showcards relates to different forms of satisficing (except in relation to response order effects – see below).

H2: By comparison, there is compelling evidence concerning the impact on response of presenting respondents with a visual stimulus compared with an auditory one. While showcards actually comprise a *mixed* visual/auditory stimulus, it seems likely that many of the conclusions relating to the effects of visual stimuli in self-completion questionnaires are relevant to face-to-face interviews using visual aids. The main form of satisficing associated with the visual presentation of response alternatives is the effect of primacy or the tendency for respondents to select items near the start of a list of alternatives in preference over later items (also known as 'response order effects' – see Krosnick and Alwin (1987). Primacy effects are assumed to arise because of the burden on short-term memory such that items near the start of the list are processed more thoroughly than items appearing later in the list. Consequently, these items are more likely to be selected by the respondent. By contrast, the opposite effect is often observed in data from telephone interviews, with respondents showing a preference for items towards the end of the list of response option (a recency effect) because these are more likely to be retained in the respondent's short-term memory. More recently, Duffy (2003) has shown that assumptions that respondents read lists of response options on showcards from top to bottom may in fact be wrong, with some respondents in his study 'learning' to pick items from lower down the list once they had answered questions where the list of options had been reversed (causing more salient/popular items to appear near the bottom of the list). Thus, there are conditions under which response order effects may be less likely to occur. Groves and Kahn (1979) also found evidence to suggest that respondents' answers are influenced by how information is displayed on the showcard: their face-to-face respondents showed greater preference for scale points that were labelled on the card compared to telephone respondents, probably because the labels draw respondents' visual attention (Groves 1990).

H3/H4: Respondents are more likely to satisfice with telephone than face-to-face interviewing. This hypothesis is supported by evidence reported by Holbrook, Green, and Krosnick (2003), who found more acquiescence, non-differentiation, and no opinion responses with telephone interviewing. Jordan, Marcus, and Reeder (1980) similarly reported more acquiescence, evasiveness, don't know and no answer responses in attitude items with telephone interviewing. Groves and Kahn (1979) reported shorter answers to open ended questions and higher missing data rates in early telephone surveys, although the differences in item non-response gradually decreased as the organisation gained experience with telephone mode.

No supporting evidence was found by De Leeuw (1992), who did not find differences in acquiescence, item non-response or the length of open-ended answers (possibly because the open questions were short and well defined).

H5: Differences in the extent of satisficing across modes are likely to be larger for low ability respondents. This hypothesis is supported by Holbrook, Green, and Krosnick (2003). Related evidence from single mode experiments is given by Narayan and Krosnick (1996), who found that in a meta analysis of over 130 studies lower education was associated with larger response order effects, acquiescence, middle alternative effects not involving status quo options, no-opinion filter effects, forbid/allow effects, balance effects and question order effects based on the norm of reciprocity.

H6: Respondents are likely to be more willing to self-disclose in a telephone interview. This is supported by evidence from Jordan et al. (1980) and Groves and Kahn (1979), who reported more extremeness in attitude items with telephone interviewing.

H7: Respondents are likely to be more willing to self-disclose in a face-to-face interview. This hypothesis is supported by Holbrook et al. (2003) and de Leeuw and van der Zouwen's (1988) meta-analysis, who found more socially desirable reporting with telephone interviewing. In other studies respondents have been found to be significantly more likely to report dentist visit during past 12 months (Weeks 1992) and less likely to admit use of illicit drugs (Aquilino 1994).

No supporting evidence was however found by de Leeuw (1992), Körmendi (1988) or Sykes and Collins (1988) or in reports of registering and turning out to vote, being arrested for drunk driving, declaring bankruptcy or having experienced any symptoms related to depression (Aneshensel et al. 1982; Groves 1979; Groves and Kahn 1979; Locander and Burton 1976).

5.4   Indicators of satisficing and their measurement

We were able to use the following indicators of respondent satisficing. Krosnik (1991) identifies additional indicators, such as endorsing the status quo instead of social change or randomly choosing response alternatives.

1. ***Completeness of responses***: Respondents who are shortcutting the response process are likely to provide less complete answers. We considered the prevalence of item non-response, measured by the proportion of 'don't know' or 'refusal' answers to all questions, and the length of answers to open-ended questions on the respondent's occupation (q25, q26), measured by the mean number of words.

2. ***Non-differentiation*** occurs when respondents choose a response category that seems appropriate for the first item of a battery of scale questions and stick to that response for all other items in the scale (Locander and Burton 1976). We measured the extent of non-differentiation using answers to four batteries of scale questions (q8a-q8g, q14-q16, q17a-q17d, q18a-q18b). For each respondent we calculated the maximum number of identical ratings made for each scale and divided it by the number of items in the scale to obtain a variable ranging from 0 to 1. We then created an overall index by averaging the scores from each of the four scales.

3. ***Acquiescence*** refers to the tendency to agree with or accept any assertion, regardless of its content (Couch and Keniston, 1976). Knowles and Condon (1999) found that acquiescence was related to shorter response latencies, supporting the view that acquiescence is a result of shortcutting rather than editing, which has been put forward as an alternative explanation. In addition, acquiescence was found to increase with the cognitive burden of a task. We used answers to agree/disagree questions (q17a-q17d, q18a, q18b) to calculate the proportion of 'agree' responses as an indicator of overall

acquiescence. At the item level we used binary indicators which took the value 1 if the respondent answered 'agree' and 0 otherwise.

4. ***Social desirability bias*** may be the result of shortcutting the retrieval stage, where the respondent echoes the socially most desirable answer without judgement about whether this corresponds to the respondent's own attitudes and instead of forming an independent judgement. It may also be the result of self-deception, an (unconscious) bias of retrieval in a way that creates a more favourable self-image of the respondent (self-deception). In either case, shortcutting can lead to retrieval of an answer that is more socially desirable than the true answer would be. We did not test the social desirability connotations of items, as suggested by Holbrook et al. (2003), but selected 21 items for which we suspect that some answer categories are likely to be more socially desirable than others (q3-q7, q9, q11-q20, q28; the selected categories are specified in the final column of table A1 in the appendix). Since we have not tested the connotations, our discussion can only refer to what we think may be potential social desirability bias. We created an overall indicator using the proportion of answers to the 21 items for which the respondent had given a socially desirable response. At the item level we created a binary indicator which took the value 1 if the response was socially desirable and 0 otherwise.

5. ***Response order effects*** are caused by differences in the sensory channel. If respondents shortcut the response process, the answers they give to closed questions will interact with the sensory channel by which the response categories are presented (Schwarz, Hippler, and Noelle-Neumann 1992). Respondents are more likely to choose one of the first answer categories if these are presented visually, and more likely to choose one of the last categories if they are presented aurally (Krosnick and Alwin 1987). We measured the extent of 'recency' and 'primacy' effects using twelve closed items (q5-q7, q11-q13, q17a-q18b), including only scale items, where the same response categories were offered for the visual and aural treatment groups. As an overall measure of primacy (recency), we calculated the proportion of answers to these twelve questions for which the first (last) category was chosen. At the item level we used binary indicators which take the value 1 if the first (last) category was chosen and 0 otherwise.

5.5 Indicators of willingness to self-disclose and their measurement

1. ***Social desirability bias*** may also be the result of deliberate editing of responses, or impression management, to make the respondent appear in a better light. The measures used are the same as those discussed for the indicators of satisficing above. If differences in the extent of social desirability bias are observed across modes, we therefore need additional clues, such as information about response latencies, to distinguish whether this is caused by deliberate editing or by shortcutting of the response process. Holtgraves (2004), for example, found that socially desirable reporting was related to longer response latencies, supporting the view that it is caused by deliberate editing.

2. ***Extreme responses***: For items where extreme categories correspond to extreme views, the extent of reporting of extremes can be seen as an indicator of the willingness to disclose sensitive information. We measured willingness to express extreme views by calculating the proportion of undesirable first- or last-category responses on the items analysed for evidence of social desirability bias. At the item level we created binary indicators which took the value 1 if the response was an extreme category which was potentially socially undesirable and 0 otherwise.

5.6 Respondents' experience of the survey interview

To shed some further light on the differences caused by modes, we also examine the background information collected about the respondent's experience of the interview. We compare the duration of interviews across modes, the respondent's willingness to continue the interview, expressed mode preferences for a hypothetical one hour survey, reported unease about answering questions on potentially sensitive topics and interviewer observations about the effort and cooperativeness of the respondents, their level of understanding of the survey questions and whether they appeared distracted or influenced by other people.

6   Results

6.1   Sample characteristics

Before turning to our analysis of mode effects, we begin by addressing the preliminary issue of how the demographic characteristics of the responding samples in the experiment compare across the two modes of interest. Table 3 shows the socio-demographic composition of the face-to-face and fixed telephone samples. (The two face-to-face groups were grouped since there were no differences in composition between the two.) The telephone sample had a significantly lower proportion of men, manual workers and respondents with low education levels. There were however no differences across modes in mean age and the proportion in work. In all subsequent analyses we adjust for these differences in sample composition, by including controls for socio-demographics in the multivariate models. The motivation for this approach was to avoid confounding of differential non-response with differential measurement error, though we recognise that some differential non-response may remain even after the socio-demographic controls.

Table 3: Sample composition

	F2F	Tele-phone	Test statistic	Degrees of freedom	P-Value
Male (%)	40.46	32.52	10.2111	$F(1, 1717)$	0.0014
Mean age (years)	56.09	55.27	0.9200	$F(1, 1702)$	0.3380
Currently in work (%)	49.27	48.98	0.0137	$F(1, 1716)$	0.9068
Manual workers (%)	36.12	25.40	19.9465	$F(1, 1641)$	0.0000
Occupation (column %)			4.7502	$F(9.93, 11540.3)$	0.0000
Senior professional	13.39	13.76			
Other professional	14.40	13.76			
Other clerical	18.46	26.52			
Senior manager or administrator	5.88	6.18			
Technical or craft	13.59	11.47			
Intermediate manual or service	8.72	5.78			
Routine manual or service	12.37	5.88			
Middle/junior manager/administrator	10.95	9.47			
Non-manual: details unknown	0.00	3.79			
Manual: details unknown	0.00	1.89			
Never had job	2.23	1.50			
High school education or less (%)	55.12	48.49	6.7900	$F(1, 1708)$	0.0092
Education (column %)			4.5937	$F(4.99, 5960.36)$	0.0004
No/basic education	11.55	6.63			
Vocational school, less than high school	16.44	10.05			
Finished high school or equivalent	26.61	31.80			
Vocational training	12.52	12.98			
College degree	14.48	17.85			
University/postgraduate degree	18.40	20.68			

Notes: F2f samples grouped, as there are no differences in composition between these modes.
Test statistics for age are from an adjusted Wald test of the equality of means and for all other variables from two-tailed Pearson χ^2 tests with the second-order correction of Rao and Scott (1984) converted into an F statistic.

Table A2 in the Appendix shows the summary statistics for the three treatment groups (unadjusted for sample composition). For comparability, the numeric items asked as open-ended questions in the aural modes were coded to correspond to the showcard categories (q1, q2, q28). For q20, the 7 showcard categories were collapsed to the 4 aural categories (see Appendix Table A1 for a summary of the differences in question format across modes).

6.2 Magnitude of mode effects

For most of the analyses we present findings from separate regressions for the three comparison groups: 1) face-to-face showcard versus telephone, 2) face-to-face showcard versus face-to-face no showcard, and 3) face-to-face versus telephone, where telephone refers to fixed line only and the estimated mode effect represents the effect of the underlined mode compared to the omitted mode. In the descriptions we sometimes refer to the face-to-face modes (regardless of showcards) and the aural modes (telephone and face-to-face without showcards). All regressions include controls for differences in sample composition: age, age squared, male, educational qualification (omitted category is university or postgraduate degree) and occupation (omitted category is senior professional).

6.2.1 Effects of mode on mean response

Table 4 summarises the main mode effects from OLS regressions of each item on mode and the controls for sample composition. For binary variables (Q9 and Q31 yes/no) logit regressions were used instead. Tables A3 to A8 in the Appendix display the results of including interactions of mode with the socio-demographics.

At the mean, mode effects in the primary comparison group (face-to-face showcard versus telephone) are visible for 13 out of 33 items. In 8 cases this seems likely to be due to the interviewer's presence, since differences are significant between the face-to-face and telephone groups regardless of whether or not showcards are used, but not between the two face-to-face groups. In 2 cases the differences seem likely to be due to the sensory channel, since differences are significant between the face-to-face groups, but not between the aural modes. In a further 3 cases the pattern is not clear.

Judging by the size of the mode coefficient compared to the standard error of the conditional mean (the constant from the regression of comparison 1), the mode effects are small. Only for Q28 (household income) is the mode coefficient larger than one standard error. This implies that differences in mean response due to mode would not lead analysts to

27

conclude that there are significant differences between groups, if they were to use for example, standard two-sample *t*-tests.

Table 4: Summary of mode effects at the mean

Question number	Question topic	Conditional mean in model (1)	Std. error of cond. mean in model (1)	Telephone vs. f2f showcard (1)	F2f no showcard vs. f2f showcard (2)	Telephone vs. f2f no showcard (3)
Q1	Time watching TV	2.0273	0.6224	-0.3629**	0.0520	-0.4508***
Q2	Time watching TV news	0.5274	0.3643	0.2896***	0.2152*	0.0834
Q3	Trust people	5.8491	0.7485	0.1451	0.0348	0.1446
Q4	Life satisfaction	9.8389	0.6365	-0.0235	0.1637	-0.1375
Q5	Political interest	3.4091	0.2641	-0.2104***	-0.0107	-0.2054***
Q6	Political understanding	2.5299	0.4017	-0.1128	0.0522	-0.1528*
Q7	Political opinion	3.3748	0.2760	0.0300	-0.0349	0.0694
Q8a	Trust institutions: parliament	4.9311	0.7633	-0.0652	-0.0011	0.0161
Q8b	Trust institutions: legal system	7.3392	0.8174	0.1401	-0.0456	0.1679
Q8c	Trust institutions: police	5.9754	0.7245	0.0387	0.1575	-0.0520
Q8d	Trust institutions: politicians	3.2024	0.6735	0.0076	-0.1295	0.1702
Q8e	Trust institutions: parties	3.5327	0.7216	-0.0222	-0.1476	0.2050
Q8f	Trust institutions: EU parliament	6.7200	0.8075	-0.0441	-0.1376	0.0738
Q8g	Trust institutions: UN	6.8128	0.8566	-0.0114	-0.2319	0.1926
Q9	Voted last national election	-0.0369	0.9149	0.0471	0.0485	0.0044
Q10	Party voted for	2.4450	0.8591	0.1359	0.2343	-0.0894
Q11	Immigration: same ethnicity	1.7926	0.2977	-0.2894***	-0.0173	-0.2674***
Q12	Immigration: different ethnicity	2.3548	0.2660	-0.2429***	-0.0656	-0.1822***
Q13	Immigration: poor outside EU	2.3507	0.2726	-0.1313*	-0.0404	-0.0908
Q14	Immigration: economy	5.5635	0.7724	0.2464	-0.0822	0.3415*
Q15	Immigration: culture	7.1121	0.7915	0.6631***	0.2443	0.4242*
Q16	Immigration: living standards	6.1133	0.7285	0.5989***	0.1285	0.5120***
Q17a	Gender role: working mothers	2.5874	0.3955	0.2777***	0.1063	0.1200
Q17b	Gender role: fathers	1.7435	0.1973	-0.1329**	0.0200	-0.1691***
Q17c	Gender role: right to jobs	4.3584	0.3918	0.1964*	0.1917*	-0.0044
Q17d	Gender role: divorce	3.5426	0.3684	-0.0550	-0.1127	0.0800
Q18a	Homosexuality	1.8741	0.4187	-0.0216	0.0344	-0.0591
Q18b	Authority	2.7838	0.2927	-0.1875***	0.0104	-0.1951***
Q19	Religiosity	4.4803	1.0020	0.3041	-0.3162	0.6708***
Q20	Church attendance	3.0887	0.2927	0.0170	0.1236*	-0.1249*
Q28	HH income	857.3306	**380.6265**	**729.4283*****	447.7694***	297.9172***
Q31	Internet access	1.4100	0.3628	0.0803	-0.0448	0.1250
Q31	Internet access: yes/no	2.1290	0.9660	0.2161	0.1045	0.0821

Notes: * p<0.05, ** p<0.01, *** p<0.001. 'Conditional mean' is the constant from the regression of responses on mode and socio-demographics (main effects only) in comparison (1); so it is the mean score in the f2f showcard group of female, non-manual workers with higher education. Mode effect refers to group underlined in the column heading. Q28 (household income) was recoded to midpoints and the final category set to 12.500. Q9 and Q31 (yes/no) were modelled using logit regression.

6.2.2 Effects of mode on the response distribution

For the ordinal items, we further tested for differences in response distributions using partial proportional odds models as outlined in Section 5.1. Note that unlike in the previous models, it was necessary to use binary indicators of low versus high education (defined as high-school level or less) and manual versus non-manual occupations instead of the full sets of indicators, in order to reduce the computational burden of the tests for proportional odds.

Tables 5A and 5B present the predicted response distributions for the face-to-face groups in the columns headed *(Col%)* and the percentage point differences in predicted probabilities due to mode in the columns headed *(% Pt Δ)*. All items that showed significant mode differences at the mean also show these differences in the ordinal model. In addition, the ordinal models identify some effects that were not visible at the mean for Q6, Q8f, Q14 and Q17d.

Mode had a linear effect on the items presented in Table 5A (the proportional odds assumption held for all items except Q11, Q12 and Q28). Telephone respondents were
- less likely to find politics too complicated (Q6),
- less likely to be against allowing immigrants of the same ethnicity (Q11), different ethnicity (Q12) or from poor countries outside Europe (Q13) to live in the country,
- more likely to say immigration enriches cultural life (Q15) and less likely to say it makes their country a worse place to live (Q16),
- less likely to (strongly) agree that mothers should cut down on paid work (Q17a) and that men have more right to jobs than women when jobs are scarce (Q17c),
- more likely to (strongly) agree that men should share responsibilities for their home and family (Q17b) and that the law should be obeyed whatever the circumstances (Q18b), and
- less likely to report a household income of less than €1000.

Mode had non-linear effects on the items in Table 5B. Telephone respondents were, for example
- more likely to watch TV for ½ to 1½ hours and less likely to watch for more than 2 hours (Q1),
- less likely to watch news on TV for less than ½ hour and more likely to watch for ½ to 1 hour (Q2),

- less likely to not at all be interested in politics (Q5),
- more likely to choose 0, 5 or 8 on an 11 point scale of their level of trust in the EU parliament (Q8f),
- less likely to choose 2 or 3 and more likely to choose 5 or 6 on an 11 point scale of how good immigration is for the economy (Q14), and
- less likely to strongly agree or strongly disagree that parents should stay together even if they do not get along (Q17d).

The conclusion drawn from the analysis at the mean that mode effects were mainly due to the interviewer's presence are robust and strengthened by the fact that for the four additional items, mode differences are also significant between each of the face-to-face groups and the telephone group, but not between the two face-to-face groups.

Table 5A: Effect of mode on response distributions: linear mode effects

		Tel-F2fSC			F2fnoSC-F2fSC		P-Value	Tel-F2fnoSC		
		F2f SC	Change	P-Value	F2f SC	Change		F2f noSC	Change	P-Value
Q6	**Never**	**19.42**	**4.36**	*	18.45	-0.65	ns	**18.67**	**4.81**	*
	Seldom	21.30	2.27		22.50	-0.45		21.17	2.58	
	Occasionally	32.44	-1.34		31.91	0.08		32.40	-1.15	
	Regularly	11.35	-1.75		13.72	0.38		12.55	-2.36	
	Frequently	15.49	-3.54		13.43	0.64		15.21	-3.88	
Q11	*Overall mode effect*			***			ns			***
	Allow many	19.81	8.36	**	21.26	-0.10		22.54	5.60	
	Allow some	30.64	0.85	**	28.50	-0.04		26.21	5.31	**
	Allow a few	33.61	0.22	***	33.80	0.06		34.10	-0.27	***
	Allow none	**15.93**	**-9.41**		16.45	0.07		**17.15**	**-10.64**	
Q12	*Overall mode effect*			***			ns			***
	Allow many	9.16	2.85		9.18	0.84		10.18	1.83	
	Allow some	22.37	3.92	*	21.95	1.22		22.68	3.59	
	Allow a few	43.66	7.50	***	44.55	-0.39		44.64	6.53	***
	Allow none	**24.81**	**-14.28**		24.32	-1.67		**22.50**	**-11.95**	
Q13	Allow many	6.07	1.90	*	6.44	0.51	ns	6.50	1.42	ns
	Allow some	16.21	3.52		17.78	0.98		16.73	2.58	
	Allow a few	46.05	0.85		42.43	0.20		47.38	0.40	
	Allow none	**31.68**	**-6.29**		33.35	-1.69		29.39	-4.41	
Q15	0 Cultural life undermined	7.10	-2.56	**	5.51	-0.67	ns	5.95	-1.57	*
	1	3.15	-0.99		4.53	-0.49		2.27	-0.56	
	2	4.69	-1.15		5.51	-0.53		4.55	-1.04	
	3	6.72	-1.70		7.72	-0.61		6.13	-1.23	
	4	6.05	-1.26		7.08	-0.43		6.16	-1.04	
	5	23.99	-2.96		21.11	-0.56		24.31	-2.13	
	6	9.00	0.25		7.51	0.07		8.63	0.06	

Item	Response	Col%	% Pt Δ	P-Value	Col%	% Pt Δ	P-Value	Col%	% Pt Δ	P-Value
	7	10.11	0.86		10.78	0.36		9.58	0.60	
	8	15.21	3.66		15.26	1.07		17.12	2.63	
	9	3.95	1.41		5.97	0.62		4.78	1.12	
	10 Cultural life enriched	**10.03**	**4.44**		9.02	1.16		**10.52**	**3.16**	
Q16	**0 Worse place to live**	**12.71**	**-4.89**	***	10.02	-0.53	ns	**11.34**	**-4.03**	***
	1	3.52	-1.16		6.01	-0.25		3.61	-1.10	
	2	7.94	-2.29		9.76	-0.33		7.88	-2.09	
	3	12.05	-2.57		11.56	-0.25		11.34	-2.25	
	4	7.28	-0.99		7.54	-0.09		7.87	-1.00	
	5	35.58	2.44		33.39	0.41		36.22	2.12	
	6	6.87	2.23		6.26	0.23		6.92	1.93	
	7	5.17	2.20		4.57	0.20		5.29	1.90	
	8	5.75	3.05		6.79	0.36		6.61	2.98	
	9	0.79	0.47		1.33	0.08		0.94	0.48	
	10 Better place to live	2.34	1.50		2.76	0.17		1.97	1.07	
Q17a	**Agree strongly**	**23.14**	**-6.63**	***	22.48	-3.26	ns	18.55	-3.35	ns
	Agree	29.77	-3.74		30.58	-1.67		29.40	-2.53	
	Neither	20.21	1.46		21.90	1.03		22.94	0.74	
	Disagree	20.42	5.89		17.22	2.34		21.38	3.25	
	Disagree strongly	6.46	3.03		7.82	1.56		7.74	1.88	
Q17b	**Agree strongly**	**74.10**	**6.67**	**	74.68	-3.00	ns	**70.95**	**9.88**	***
	Agree	20.51	-5.00		17.40	1.80		22.28	-7.18	
	Neither	3.03	-0.92		5.84	0.85		5.08	-2.10	
	Disagree	1.92	-0.60		1.43	0.22		1.34	-0.49	
	Disagree strongly	0.44	-0.14		0.66	0.11		0.35	-0.12	
Q17c	**Agree strongly**	**27.14**	**-5.98**	*	**27.37**	**-4.16**	*	22.20	-1.79	ns
	Agree	17.25	-1.65		18.15	-1.18		15.65	-0.56	
	Neither	22.29	0.40		21.96	0.35		25.08	0.11	
	Disagree	23.96	4.16		20.92	2.47		25.84	1.27	
	Disagree strongly	9.36	3.07		11.59	2.53		11.23	0.96	
Q18b	**Agree strongly**	**61.24**	**12.17**	***	59.78	-1.75	ns	**59.22**	**13.95**	***
	Agree	23.04	-6.06		27.32	0.98		24.84	-7.15	
	Neither	11.46	-4.30		9.79	0.56		12.28	-5.10	
	Disagree	3.39	-1.44		2.52	0.16		2.96	-1.37	
Q28	*Overall mode effect*		***			***				ns
	< €1000	**77.57**	**-46.25**	***	**77.64**	**-47.65**	***	30.16	1.52	
	€1000-1500	15.23	8.83	***	15.20	13.94	***	28.86	-4.49	
	€1500-2000	3.68	5.57	***	3.64	8.52	***	11.29	-1.98	
	€2000-2500	1.46	16.43	***	1.46	15.65	***	17.42	-0.15	
	€2500-3000	1.17	7.04	***	1.18	5.04	**	6.54	1.46	
	€3000 or more	0.88	8.40		0.88	4.51		5.73	3.64	

Notes: The proportional odds assumption holds in all comparisons except for items Q11, Q12 and Q28. (Col%) shows the predicted response distribution based on partial proportional odds models estimated with Stata's gologit2 command (Williams 2006). (% Pt Δ) shows the percentage point difference in predicted probabilities between modes. (P-Value) reports the significance of mode effects. For all items except Q11, Q12 and Q28, the proportional odds assumption holds for all comparisons. In these models a single mode coefficient is estimated. For the remainder items the significance of mode effects is reported a) overall, based on Wald tests of the joint hypothesis that the mode coefficients in each of a series of binary logistic regressions of Probability(Y>j) equal zero, and b) separately for each binary equation. * p<0.05, ** p<0.01, *** p<0.001.

Table 5B: Effect of mode on response distributions: non-linear mode effects

		F2f SC (Col%)	Tel-F2fSC (% Pt Δ)	Tel-F2fSC (P-Value)	F2f SC (Col%)	F2fnoSC-F2fSC (% Pt Δ)	F2fnoSC-F2fSC (P-Value)	F2f noSC (Col%)	Tel-F2fnoSC (% Pt Δ)	Tel-F2fnoSC (P-Value)
Ql	*Overall mode effect*			***			*			***
	0 hrs	2.59	0.22		2.49	-0.21		1.89	1.12	***
	0 - 1/2 hr	5.43	-3.52	*	5.25	**-3.80**	**	1.29	0.70	***
	1/2 - 1 hr	12.73	**13.14**	***	12.57	**4.36**		18.63	**6.47**	***
	1 - 1 1/2 hrs	7.42	1.24	***	7.48	-0.92		7.14	1.26	***
	1 1/2 - 2 hrs	21.04	-0.11	***	21.51	0.46		20.55	1.07	***
	2 - 2 1/2 hrs	6.81	-0.43	**	7.01	-2.95		5.80	-0.30	***
	2 1/2 - 3 hrs	18.80	**-6.78**		19.02	2.33		16.52	-2.17	***
	> 3 hrs	25.17	-3.75		24.68	0.73		28.17	**-8.15**	
Q2	*Overall mode effect*			***			***			ns
	0 hrs	4.62	-0.10		4.49	-0.32		4.58	-0.19	
	0 - 1/2 hr	32.98	**-22.69**	***	33.14	**-24.21**	***	10.11	-0.41	
	1/2 - 1 hr	36.99	**21.50**		37.28	**26.69**		60.79	-0.93	
	1 - 1 1/2 hrs	11.19	-2.03		11.09	-4.18		8.16	0.33	
	1 1/2 - 2 hrs	7.26	3.98		7.15	4.28		10.90	0.71	
	2 - 2 1/2 hrs	1.53	0.13		1.50	-1.30		1.13	0.09	
	2 1/2 - 3 hrs	2.61	0.38		2.56	0.13		2.76	0.24	
	> 3 hrs	2.82	-1.18		2.77	-1.10		1.57	0.15	
Q5	*Overall mode effect*			***			ns			***
	Very interested	12.60	**5.87**	***	16.16	0.21		17.66	-0.22	***
	Quite interested	39.63	4.99	***	34.92	0.40		32.82	**13.22**	***
	Hardly interested	29.34	-4.78	***	28.16	-0.26		29.20	-3.83	
	Not at all interested	18.42	**-6.06**		20.76	-0.35		20.32	**-9.18**	
Q8f	*Overall mode effect*			**			ns			**
	0 No trust at all	4.86	**4.11**	*	4.56	0.69		4.92	**4.11**	*
	1	4.60	-3.43		3.86	0.53		3.62	-2.44	
	2	7.96	-2.24		6.91	0.80		6.67	-0.92	
	3	8.74	-0.07		9.55	0.84		11.26	-2.57	
	4	9.98	-2.11		10.46	0.57		11.61	-3.74	
	5	19.06	**6.63**		21.14	0.13		23.43	2.27	
	6	12.79	-0.83		13.42	-0.58		13.37	-1.44	
	7	11.63	-2.49		10.09	-0.74		7.74	1.35	
	8	11.88	3.07		11.76	-1.19		10.43	**4.47**	
	9	4.99	-2.83		5.70	-0.70		5.71	-3.55	*
	10 Complete trust	3.51	0.18		2.54	-0.34		1.24	2.45	
Q14	*Overall mode effect*			***			ns			***
	0 Bad for economy	10.86	1.53		9.94	0.57		9.98	2.40	
	1	7.89	**-6.22**		6.79	0.31		6.05	-4.35	
	2	7.15	-3.14	*	9.81	0.34		12.85	**-8.83**	***
	3	9.14	1.14		9.87	0.21		10.66	-0.37	***
	4	9.65	-0.88		8.47	0.09		7.25	1.52	**
	5	28.96	3.34		27.92	-0.30		26.08	**6.12**	
	6	5.05	**4.34**		6.48	-0.20		7.50	2.02	
	7	7.27	0.73		7.09	-0.28		6.89	0.98	
	8	6.85	1.47		7.00	-0.35		6.86	1.52	
	9	2.22	-0.60		2.50	-0.14		2.79	-1.25	
	10 Good for economy	4.96	-1.72		4.13	-0.25		3.09	0.24	
Q17d	*Overall mode effect*			***			ns			**

	Col%	% Pt Δ	P-Value	Col%	% Pt Δ	P-Value	Col%	% Pt Δ	P-Value
Agree strongly	10.42	-4.42	*	9.50	1.79		10.47	**-4.55**	**
Agree	12.01	-1.07	*	10.45	1.56		10.63	0.27	
Neither	19.15	5.08		21.72	1.76		25.97	-1.67	
Disagree	28.78	**10.94**	***	30.19	-0.89		30.49	**9.33**	
Disagree strongly	29.65	**-10.54**		28.14	-4.22		22.44	-3.39	
Q19 *Overall mode effect*			ns			ns			***
0 Not at all religious	24.19	-2.58		24.17	4.13		27.68	**-6.37**	*
1	3.64	-0.24		5.17	0.49		5.68	-3.05	**
2	5.35	-0.29		6.58	0.44		7.75	-2.96	***
3	7.64	-0.27		7.59	0.27		9.29	-1.20	***
4	5.06	-0.09		5.10	0.05		3.93	0.41	***
5	20.32	0.37		17.01	-0.60		16.04	**6.04**	*
6	5.83	0.31		4.62	-0.36		4.51	2.36	
7	7.62	0.55		7.70	-0.79		7.73	0.85	
8	8.98	0.85		6.95	-0.92		4.26	5.74	
9	2.59	0.28		3.83	-0.59		3.69	-1.19	
10 Very religious	8.79	1.08		11.28	-2.11		9.44	-0.62	

Notes: (Col%) shows the predicted response distribution based on partial proportional odds models estimated with Stata's gologit2 command (Williams 2006). (% Pt Δ) shows the percentage point difference in predicted probabilities between modes. (P-Value) reports the significance of mode effects a) overall, based on Wald tests of the joint hypothesis that the mode coefficients in each of a series of binary logistic regressions of Probability(Y>j) equal zero, and b) separately for each binary equation. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

6.2.3 Impact of mode effect on relationship between variables

Although the mode of data collection may affect response distributions, it may not necessarily affect the relationships between variables. Therefore we tested the effect of mode on summed attitude scores and on the relationship between summed scores and other variables using OLS regressions. We examined the relationships between summed scores and items where we did and did not find mode effects. The rationale behind this was that if the summed score and the independent item are both affected by mode in the same way, then the relationship between the two might not be changed. However, if one of the two is affected by mode, but not the other, then the relationship might be altered by the mode of data collection.

Table 6: Effect of mode on summed scores and relationships between variables

			Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
			Coeff.	P-Value (std. err.)	Coeff.	P-Value (std. err.)	Coeff.	P-Value (std. err.)
Q5-7: Political interest	(1)	Cond. Mean	8.6173***	(0.6601)	7.6210***	(0.7220)	9.3008***	(0.6549)
		Mode	-0.3543**	0.1356	0.0818	0.1458	-0.4316**	0.1355
	(2)	Mode	-0.4249	0.3222	0.0384	0.9293	-0.4416	0.2970
		Voted	-0.9823**	0.0035	-1.0009**	0.0026	-0.8839**	0.0071
		Voted*mode	0.0749	0.8675	0.0154	0.9732	0.0312	0.9439
	(3)	Mode	-0.4764	0.0627	0.0072	0.9803	-0.5424	0.0612
		TV news watching	-0.3266***	0.0000	-0.3509***	0.0000	-0.3142***	0.0004
		TV News*mode	0.0792	0.3773	0.0339	0.7578	0.0731	0.5020
Q8a-g: Trust in institutions	(1)	Cond. Mean	39.7383***	(4.6739)	31.0810***	(4.8119)	42.1856***	(4.7274)
		Mode	0.0117	0.9858	-0.9767	0.9844	1.0480	0.9448
	(2)	Mode	-3.2120	0.1525	-3.9128	0.0501	1.0662	0.6204
		Trust in people	1.1539***	0.0001	1.1557***	0.0000	1.8609***	0.0000
		Trust peop.*mode	0.6597	0.1074	0.6749	0.0756	-0.0699	0.8596
	(3)	Mode	-6.7273	0.0613	-5.3460	0.1285	-0.4035	0.9007
		Political interest	-1.4100***	0.0001	-1.2843***	0.0004	-0.6714*	0.0205
		Pol. interest*mode	0.8885	0.0567	0.5803	0.1932	0.1741	0.6780
Q11-16: Immigration	(1)	Cond. Mean	26.7354***	(2.5738)	29.1639***	(2.8010)	28.5484***	(2.5901)
		Mode	2.4619***	0.5598	0.4409	0.6030	2.0815***	0.5550
	(2)	Mode	2.7213*	0.0155	(dropped)		(dropped)	
		Party1 voted for	0.6088	0.8615	1.6784	0.6361	-1.8355	0.6230
		Party1*mode	1.4684	0.6890	-3.6461	0.4723	3.9389	0.3177
		Party2 voted for	0.7926	0.4803	0.6555	0.5541	0.5653	0.5984
		Party2*mode	-1.5976	0.2912	0.0474	0.9751	-1.4440	0.3217
		Party3 voted for	8.8354***	0.0000	6.7542***	0.0002	8.8448***	0.0001
		Party3*mode	1.0832	0.7315	(dropped)		(dropped)	
		Party4 voted for	-1.3494	0.8200	-2.1578	0.7303	-7.7307	0.0856
		Party4*mode	-0.5807	0.9411	-6.4414	0.4041	4.2888	0.5216
		Party5 voted for	-2.1104	0.4023	-1.5339	0.6670	1.8436	0.6609
		Party5*mode	-3.9422	0.1736	3.2865	0.4788	-6.8727	0.1280
		Party6 voted for	4.1224*	0.0161	3.9955*	0.0218	6.4850***	0.0000
		Party6*mode	-1.8492	0.3664	3.3195	0.1568	-4.2536*	0.0251
	(3)	Mode	1.9800*	0.0127	0.8990	0.2786	1.1372	0.1278
		In work	-0.8995	0.3456	-0.6156	0.5138	-2.1240*	0.0226
		In work*mode	0.8834	0.4173	-0.9122	0.4448	1.8290	0.0889
Q17a-d: Gender roles	(1)	Cond. Mean	14.7002***	(0.7690)	14.1401***	(0.8800)	14.5312***	(0.7971)
		Mode	0.6061***	0.1701	0.2086	0.1821	0.3829*	0.1602
	(2)	Mode	0.6224**	0.0073	0.1687	0.5046	0.4345*	0.0478
		In work	0.3252	0.2733	0.2472	0.4208	0.5639*	0.0350
		In work*mode	-0.0075	0.9819	0.0946	0.7914	-0.0928	0.7692
	(3)	Mode	0.5030	0.0629	-0.0058	0.9835	0.4259	0.0921
		Religiosity	-0.1787***	0.0000	-0.1835***	0.0000	-0.1582***	0.0000
		Religiosity*mode	0.0322	0.5243	0.0347	0.5201	0.0135	0.7815

Notes: * p<0.05, ** p<0.01, *** p<0.001. Model (1) regresses the summed score on mode and socio-demographics; models (2) and (3) each include an additional explanatory variable plus its interaction with mode. The analysis on includes cases with valid answers for all items included in scale. For Q11-16 the omitted party as well as parties 1, 5, 6 and 7 are socialist-liberal; parties 2, 3 and 4 are centre right. Although Q11-Q13 and Q14-Q16 have different numbers of response categories, the results are comparable if analysed separately.

Table 6 indicates that the mode effects in the summed scores mirror the item level mode effects. The observed mode effects are relatively small (the coefficient is less than one standard error of the conditional mean summed score), and differences again appear to be due to the presence of the interviewer rather than to the effects of using showcards. However, mode does not appear to affect relationships between scores and other scores or variables (that is, the interactions between mode and the independent variable are not significant), even if the summed scores showed mode effects but the independent variable did not, and vice versa. We also tested the effect of mode on the relationship between response distributions and other variables using proportional odds models (Appendix table A11). The results from this specification should however be interpreted with caution since we did not allow for the non-proportional effects of mode on response distributions found in table 5, but instead constrained all mode effects to be proportional. Nonetheless, the results suggest that mode does not affect the relationship between response distribution and other variables (except for time spent watching news programmes on television, where there is a positive interaction between telephone respondents and political interest).

6.3   Nature of mode effects

In the following we examine the nature of mode differences and test the hypotheses about the potential causes of mode effects, both overall and at the level of individual items. Although the hypotheses and discussions of findings focus on particular comparisons of modes, most tables include all comparison groups.

6.3.1   Item non-response

Unlike previous studies, we did not find that item non-response rates overall are higher with telephone mode; in fact we found the opposite, that item non-response (refusal) rates were higher with face-to-face (no showcards) than with telephone. Item non-response was mainly driven by non-response to the household income question. For this the overall missing rate is again higher with face-to-face no showcard although the proportion of 'don't know' answers is higher with telephone than with the face-to-face no showcard group.

These observations do not lend support to H3 and H4, according to which we would expect to see more incomplete answers with telephone interviewing than with f2f showcards. Comparing the two f2f groups, there are no overall differences, although refusal rates for the income question are significantly higher without the showcards. This may be seen as weak

support for H1, according to which one would expect to see more complete answers with the showcards than without, though it may alternatively be an effect of increased anonymity of response with the showcard (as the respondent need only read out a code letter rather than say an actual money amount).

Table 7: Mean item non-response rate by mode (%)

	F2f showcard	F2f no showcard	Telephone	Telephone vs. f2f showcard	F2f no showcard vs. f2f showcard	Telephone vs. f2f no showcard
Overall:						
Don't know	2.56	2.75	2.41	ns	Ns	ns
Refusal	1.79	1.95	1.75	ns	Ns	*
Total item non-response	4.35	4.70	4.17	ns	Ns	*
Income:						
Don't know	2.72	3.47	6.91	***	Ns	*
Refusal	20.39	32.05	16.85	*	***	***
Total item non-response	23.11	35.52	23.76	ns	***	***

Notes: * p<0.05, ** p<0.01, *** p<0.001, ns=not significant. Asterisks denote P-values of mode main effects from logit regressions of indictors of item non-response on mode and demographics.

6.3.2 Response to open-ended questions

For the length of answers to open-ended questions we find evidence that supports H3/H4, according to which we expect shorter answers with telephone compared to the face-to-face modes. The difference is significant for the description of occupation but not for the occupational title, which may be expected since the latter tends to be short in any case. A caveat about this measure, however, is that differences in reporting by the respondent are confounded with way interviewers record the verbatim answers. Telephone interviewers may be more tempted to abbreviate responses to shorten awkward silences.

Table 8: Length of response to open-ended questions by mode

	F2f Mean	Telephone Mean	Telephone vs. f2f Coeff.	P-Value
q25 (occupation title)	1.47	1.50	-0.0393	0.3379
q26 (occupation description)	2.15	1.71	-0.6052***	0.0000

Notes: Unconditional mean number of words and coefficients and P-values of mode main effects from regressions of word counts on mode and demographics where face-to-face is the omitted category. The two face-to-face groups were combined, since the questions did not involve showcards for either group.

6.3.3 Non-differentiation

We do not find evidence for H1 or H3/H4 overall or at the item level, according to which we would expect more non-differentiation without showcards than with showcards and more non-differentiation with telephone interviewing than in the face-to-face no showcard mode. We do not find evidence to support H5, that mode effects are likely to be bigger for low ability respondents.

There are however some differences between the telephone and face-to-face showcard groups, suggesting that although showcards and interviewer presence do not lead to differences on their own, their interaction may.

Table 9: Percentage of non-differentiated answers

	F2f showcards	F2f no showcards	Telephone
Proportion of non-differentiated answers	54.93	54.01	53.35

Notes: Significance of differences between modes tested by regressing the proportion of non-differentiated answers on mode and demographics. The only significant difference is between F2f showcards and telephone, where for the main mode effect the P-Value=0.004. None of the interactions between mode and demographics were significant.

Table 10: Percentage of non-differentiated answers per scale

	Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q8a – q8g	-0.0068	0.5218	0.0014	0.8962	-0.0074	0.4651
q14, q15, q16	-0.0032	0.8356	-0.0067	0.6827	0.0070	0.6516
q17a – q17d	-0.0359**	0.0011	-0.0213	0.0683	-0.0131	0.2026
q18a, q18b	-0.0301*	0.0277	-0.0095	0.4989	-0.0144	0.2719

Notes: * p<0.05, ** p<0.01, *** p<0.001. Coefficients and P-values of main mode effects from regressions of the proportion of non-differentiated answers to a battery of questions on mode and demographics. None of the interactions with socio-demographics were significant.

6.3.4 Acquiescence

We again do not find evidence to support H1 or H3/4 or H5. In fact overall, and for item q18b, we find the opposite: more acquiescence with face-to-face no showcard than telephone interviewing.

Table 11: Percentage 'agree'

	F2f showcard	F2f no showcard	Telephone
Proportion agree/strongly agree	21.10	21.69	18.97

Notes: P-Value of main mode effect is 0.014 for f2f no showcard vs. telephone in regressions of the proportion of 'agree' answers on mode and demographics. None of the remainder main mode effects or interactions with demographics were significant.

Table 12: Whether 'agree' by item

	Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
17a	-0.0901	0.5264	0.1126	0.4329	-0.1509	0.2760
17a	1.3556	0.3500	3.2910*	0.0276	-2.0148	0.1295
age*mode	ns	ns	-0.1156*	ns	ns	0.2956
17b	-0.0637	0.7094	0.0807	0.6303	-0.1407	0.3938
17c	-0.1965	0.2451	-0.0869	0.6116	-0.0903	0.6055
17d	0.0587	0.7784	-0.1146	0.5946	0.0985	0.6441
18a	0.2553	0.0791	0.1038	0.4895	0.1180	0.4012
18a	-0.6497	0.6366	0.6739	0.6458	-1.2394	0.3469
lowedu*mode	0.7671*	ns	ns	ns	ns	ns
18b	-0.6709***	0.0000	0.1513	0.2969	-0.8118***	0.0000
18b	0.2964	0.8395	-0.5093	0.7333	1.0142	0.4902
man*mode	ns	ns	0.6331*	ns	ns	ns

Notes: * p<0.05, ** p<0.01, *** p<0.001, ns=not significant. Coefficients and P-Values for main mode effects from logit regressions of whether the respondent answered 'agree' on mode and demographics and significance levels of interactions of demographic variables with mode.

6.3.5 Response order effects

We find hardly any support for H2, according to which we should see more first category or 'primacy' responses in face-to-face interviews with showcards and more last category or 'recency' responses from respondents interviewed without showcards. Overall, there are no differences between the two face-to-face groups. At the item level, there are only differences in the expected direction for primacy (Q17a, Q17c). But if this was a response order effect due to showcards one would not expect to see any differences between the telephone and face-to-face no showcard groups.

Table 13: Response order effects

	F2f showcard	F2f no showcard	Telephone
Recency: last answer category (%)	15.22	14.13	10.60
Primacy: first answer category (%)	25.60	24.57	23.06

Notes: P=0.000 for the main mode effect between f2f showcard vs. telephone and f2f no showcard vs. telephone, in regressions of the proportion of answers from the last answer category on mode and demographics (recency). P=0.002 for primacy regressions between f2f no showcard and telephone. All remainder main and interaction effects non-significant.

Table 14: Recency by item

	Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q6	-0.0389	0.8389	-0.0609	0.7550	0.0532	0.7849
q6	-0.3579	0.8638	-0.0434	0.9832	-0.2866	0.8887
manual*mode	0.8605*	ns	ns	ns	0.8625*	ns
q7	-0.0871	0.6269	-0.1573	0.3822	0.0675	0.7085
q11	-1.0867***	0.0000	0.0573	0.7521	-1.0736***	0.0000
q12	-1.0963***	0.0000	-0.1748	0.2770	-0.8530***	0.0000
q12	0.8859	0.5741	-1.4545	0.3562	2.5247	0.1334
age*mode	ns	ns	ns	ns	-0.1328*	ns
age2*mode	ns	ns	ns	ns	ns	0.0013*
q13	-0.5375***	0.0002	-0.1687	0.2407	-0.3120*	0.0303
q13	0.7325	0.5934	-2.2412	0.1474	3.1603*	0.0290
age*mode	ns	ns	ns	ns	-0.1513**	ns
age2*mode	ns	ns	-0.0011*	ns	0.0015**	ns
q17a	0.0545	0.8104	0.1334	0.5786	-0.2368	0.2786
q17a	1.5267	0.5738	-0.0691	0.9808	0.9807	0.6997
manual*mode	-1.6304**	ns	ns	ns	ns	ns
q17b	-2.0892*	0.0415	-0.6905	0.4749	-2.7473	0.0606
q17c	-0.1697	0.4050	0.0899	0.6513	-0.3294	0.1027
q17d	-0.6309***	0.0001	-0.4018**	0.0090	-0.2909	0.0681
q18a	-0.2628	0.1483	-0.1270	0.4784	-0.1194	0.5132
q18b	-1.3690	.	-0.4917	0.5583	-0.4151	0.6273
q18b	5.4841	0.5059	-1.3784	0.7290	7.8184	.
manual*mode	-17.4218***	ns	ns	ns	ns	ns

Notes: * p<0.05, ** p<0.01, *** p<0.001, ns=not significant.

Table 15: Primacy by item

	Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q5	-1.4110	0.5331	1.3287	0.5541	-2.6227	0.2502
q6	0.1730	0.2987	-0.1436	0.4022	0.3494*	0.0413
q6	-2.1812	0.1827	-5.0661*	0.0209	3.1356	0.1322
age*mode	ns	ns	0.2075*	ns	ns	ns
age2*mode	-0.0020*	ns	ns	ns	ns	ns
q7	-0.9591*	0.0272	-0.1038	0.7774	-0.8920*	0.0300
q7	0.6791	0.8660	8.3472*	0.0319	-7.2363*	0.0209
age*mode	ns	ns	-0.3367**	ns	0.2903*	ns
age2*mode	0.0029**	ns	ns	ns	ns	ns
manual*mode	2.0376*	ns	ns	ns	ns	ns
q11	0.4792**	0.0021	0.1383	0.4079	0.3369*	0.0271
q11	-3.1125*	0.0270	-3.3520*	0.0463	-0.3741	0.8121
age*mode	0.1178*	ns	ns	ns	ns	ns
q12	0.3299	0.1329	0.0972	0.6703	0.2896	0.1811
q13	0.2029	0.4211	0.0494	0.8538	0.1749	0.4917
q17a	-0.5717***	0.0004	-0.3738*	0.0199	-0.1288	0.4355
q17b	0.3503*	0.0226	-0.1170	0.4246	0.4907***	0.0010
q17c	-0.4273**	0.0062	-0.4181**	0.0076	0.0013	0.9936
q17d	-0.5322*	0.0212	-0.0258	0.9069	-0.6456**	0.0052
q17d	1.2352	0.6258	1.2581	0.5799	-1.0389	0.6721

39

man*mode	ns	ns	-0.9021*	ns	ns	ns
q18a	-0.2121	0.2042	-0.2133	0.2088	0.0216	0.8980
q18b	0.6933***	0.0000	-0.1120	0.4024	0.8074***	0.0000
q18b	-0.7301	0.5813	0.8333	0.5448	-1.7597	0.1907
lowedu*mode	ns	ns	ns	ns	0.6493*	ns

Notes: * p<0.05, ** p<0.01, *** p<0.001, ns=not significant.

There are however differences in the likelihood of choosing extreme responses between telephone and both face-to-face groups, although they do not appear to be explained by the use of showcards: telephone respondents are less likely to choose 'recent' answers than either face-to-face group. The direction is not clear for primacy.

When the second half of the response categories was compared to the first half, instead of including only the first and last response category, there were no differences between modes (table not shown).

6.3.6   Social desirability bias

Overall, there is evidence to support H7, according to which we expect more potentially desirable answers with telephone interviewing than face-to-face no showcards. At the item level there are differences for about half the items for which there were significant mode effects in the initial analysis of the magnitude of mode effects comparing face-to-face showcard and telephone.

Table 16: Overall percentage of socially desirable answers

	F2f showcard	F2f no showcard	Fixed telephone
Percentage	40.15	40.15	44.41

Table 17: Potential social desirability bias by item

		Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
		Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
	Overall	0.0337***	0.0001	0.0018	0.8360	0.0337***	0.0001
	Overall	(dropped)	–	–	–	–	–
	man*mode	0.0358*	0.0460	–	–	–	–
	age2*mode	-0.0001*	0.0294	–	–	–	–
M	q1	0.1097***	0.0001	-0.0031	0.9127	0.1173***	0.0000
	q1	(dropped)	–	–	–	–	–
	man*mode	0.1263*	0.0388	–	–	–	–
M	q2	-0.0487*	0.0458	–	–	–	–
	q2	(dropped)	–	–	–	0.2172	0.3420
	manual*mode	-0.1467**	0.0093	–	–	-0.1426*	0.0156
	q3	-0.0205	0.4121	0.0155	0.5487	-0.0293	0.2345
	q3	–	–	-0.5283*	0.0324	–	–
	age*mode	–	–	0.0190*	0.0367	–	–

40

	age2*mode	−	−	-0.0002*	0.0458	−	−
	q4	-0.0166	0.5679	0.0278	0.3486	-0.0375	0.1893
M	q5	0.0992***	0.0010	-0.0169	0.5885	0.1163***	0.0001
	q5	(dropped)	−	−	−	-0.6125*	0.0359
	age*mode	0.0228*	0.0356	−	−	0.0280*	0.0112
	age2*mode	−	−	−	−	-0.0003**	0.0099
	q6	0.0367	0.2148	-0.0257	0.3883	0.0557	0.0527
	q6	(dropped)	−	−	−	-0.0119	0.9668
	manual*mode	-0.2018**	0.0044	−	−	-0.1786**	0.0085
	q7	0.0114	0.7018	-0.0275	0.3674	0.0462	0.1100
	q9	0.0084	0.6871	0.0060	0.7886	0.0064	0.7531
M	q11	0.0897**	0.0031	-0.0136	0.6600	0.1025***	0.0005
	q11	(dropped)	−	-0.5814	0.0624	−	−
	age*mode	0.0253*	0.0206	−	−	−	−
	age2*mode	-0.0002*	0.0163	-0.0002*	0.0463	−	−
M	q12	0.0568	0.0527	0.0150	0.6073	0.0459	0.1096
	q12	(dropped)	−	−	−	−	−
	man*mode	0.1837**	0.0034	−	−	−	−
M	q13	0.0123	0.6480	0.0004	0.9874	0.0158	0.5464
	q14	0.0098	0.7010	-0.0184	0.4567	0.0319	0.1833
M	q15	0.0699*	0.0219	0.0122	0.6927	0.0603*	0.0442
M	q16	0.0555*	0.0179	0.0111	0.6132	0.0438	0.0586
M	q17a	-0.1061***	0.0006	-0.0374	0.2400	-0.0478	0.1170
	q17a	(dropped)	−	0.8001*	0.0114	−	−
	age*mode	-0.0215*	0.0470	-0.0323**	0.0068	−	−
	age2*mode	−	−	0.0003*	0.0144	−	−
	manual*mode	−	−	−	−	-0.2020**	0.0052
M	q17b	0.0505***	0.0004	-0.0108	0.5533	0.0660***	0.0000
	q17b	−	−	−	−	0.1044	0.5439
	lowedu*mode	−	−	−	−	0.0682*	0.0328
M	q17c	0.0384	0.1883	0.0102	0.7258	0.0242	0.4037
	q17d	-0.0358	0.1306	-0.0143	0.5730	-0.0350	0.1364
	q18a	0.0177	0.5644	-0.0098	0.7500	0.0239	0.4253
	q18a	(dropped)	−	−	−	-0.3916	0.1857
	lowedu*mode	0.1428*	0.0376	−	−	0.1719*	0.0111
M	q18b	0.0354	0.0859	0.0032	0.8849	0.0311	0.1158
	q20	0.0071	0.7119	-0.0352	0.0504	0.0455**	0.0085
	q19	0.0271	0.3505	-0.0376	0.1930	0.0680*	0.0163
M	q28	0.2613***	0.0000	0.1844***	0.0000	0.0735**	0.0050
	q28	(dropped)	−	0.0840	0.7315	-0.1183	0.6702
	man*mode	−	−	0.1743***	0.0001	−	−
	age*mode	0.0226**	0.0061	−	−	−	−
	age2*mode	-0.0003***	0.0002	−	−	-0.0002*	0.0179
	manual*mode	-0.1043*	0.0224	-0.0899*	0.0395	−	−
	lowedu*mode	−	−	−	−	-0.1171*	0.0444

Notes: M in first column indicate items for which there were mode effects between F2f showcard and telephone at the mean. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.
Coefficients and P-Values of the main mode effect in logit regressions of whether an answer is socially desirable on mode and demographics. Coefficients and asterisks given for interaction effects indicate significant interactions in models including all main and interaction effects.

6.3.7 Extreme responses

Differences between groups were observed in the propensity to select first- and last-category responses. Contrary to our hypotheses, these differences could not be attributed to the use of a visual stimulus in the face-to-face with showcard condition, since there were no differences observed between the two face-to-face groups. We therefore examined whether differences in reporting of extremes are related to social desirability concerns, by testing the hypothesis that the mode that led to more social desirability bias also leads to less reporting of socially undesirable extremes.

Consistent with the finding that telephone respondents were more likely to give socially desirable responses, they were also less likely to select socially undesirable extremes. At the item level, mode did not have an effect for 10 of the items tested, but did lead to less reporting of extremes compared to both face-to-face groups for 6 of the items (tables 18 and 19).

Table 18: Overall percentage of extreme socially undesirable responses

	F2f showcard	F2f no showcard	Fixed telephone
Percentage	15.96	14.90	11.93

Table 19: Extreme socially undesirable responses

	Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
	Coeff.	P-Value	Coeff.	Coeff.	P-Value	Coeff.
Overall	-0.0361***	0.0000	-0.0132	0.0850	-0.0251***	0.0001
q3	0.0536	0.7993	-0.1277	0.5497	0.2499	0.2442
q3	-2.0231	0.3572	-0.1290	0.9522	-2.2993	0.3416
manual*mode	0.9790*	0.0357	1.2553*	0.0122	-0.1508	0.7312
q4	-0.1739	0.5608	-0.4093	0.2113	0.2106	0.5183
q4	9.9670*	0.0360	10.7122*	0.0326	-0.0606	0.9896
man*mode	-0.9756	0.1580	-1.6662*	0.0349	0.6391	0.4330
age*mode	-0.3182*	0.0415	-0.3101	0.0818	-0.0257	0.8710
q5	-0.6431***	0.0003	-0.0119	0.9415	-0.7015***	0.0001
q6	-0.0389	0.8389	-0.0609	0.7550	0.0532	0.7849
manual*mode	0.8605*	0.0466	-0.0399	0.9318	0.8625*	0.0406
q7	-0.9591*	0.0272	-0.1038	0.7774	-0.8920*	0.0300
q7	0.6791	0.8660	8.3472*	0.0319	-7.2363*	0.0209
age*mode	-0.0310	0.8388	-0.3367**	0.0065	0.2903*	0.0272
age2*mode	0.0002	0.8898	0.0029**	0.0072	-0.0026	0.0521
manual*mode	1.1058	0.2517	2.0376*	0.0143	-1.0909	0.2076
q11	-1.0867***	0.0000	0.0573	0.7521	-1.0736***	0.0000
q12	-1.0963***	0.0000	-0.1748	0.2770	-0.8530***	0.0000
q12	0.8859	0.5741	-1.4545	0.3562	2.5247	0.1334
age*mode	-0.0830	0.1621	0.0446	0.4456	-0.1328*	0.0355

age2*mode	0.0008	0.1406	-0.0004	0.4027	0.0013*	0.0248
q13	-0.5375***	0.0002	-0.1687	0.2407	-0.3120*	0.0303
q13	0.7325	0.5934	-2.2412	0.1474	3.1603*	0.0290
age*mode	-0.0411	0.4146	0.1058	0.0634	-0.1513**	0.0052
age2*mode	0.0004	0.4209	-0.0011*	0.0293	0.0015**	0.0021
q14	0.1898	0.3774	-0.1285	0.5666	0.4118	0.0591
q15	-0.1037	0.7143	-0.3337	0.2839	0.2793	0.3664
q15	-4.0631	0.1081	-7.0056	0.0520	2.4699	0.4897
man*mode	-0.8945	0.1420	-1.7909**	0.0094	0.9520	0.1789
age*mode	0.1681	0.0634	0.3336*	0.0173	-0.1458	0.2911
age2*mode	-0.0013	0.1050	-0.0034*	0.0134	0.0019	0.1610
q16	-0.2575	0.2614	-0.3368	0.1548	0.0610	0.8006
q17a	0.0545	0.8104	0.1334	0.5786	-0.2368	0.2786
q17a	1.5267	0.5738	-0.0691	0.9808	0.9807	0.6997
manual*mode	-1.6304**	0.0086	-0.4446	0.4349	-1.0354	0.0965
q17b	-2.0892*	0.0415	-0.6905	0.4749	-2.7473	0.0606
q17c	-0.4273**	0.0062	-0.4181**	0.0076	0.0013	0.9936
q17d	-0.6309***	0.0001	-0.4018**	0.0090	-0.2909	0.0681
q18a	-0.2628	0.1483	-0.1270	0.4784	-0.1194	0.5132
q18b	-1.3690	.	-0.4917	0.5583	-0.4151	0.6273
q18b	5.4841	0.5059	-1.3784	0.7290	7.8184	.
manual*mode	-17.4218***	0.0000	0.2710	0.8757	-17.9829	.
q19	-0.2533	0.0999	0.0623	0.6797	-0.3609*	0.0147
q20	0.0890	0.4885	0.1741	0.1878	-0.1245	0.3241

Notes: * p<0.05, ** p<0.01, *** p<0.001.
Coefficients and P-Values of the main mode effect in logit regressions of whether an answer is an extreme socially undesirable category on mode and demographics. Coefficients and asterisks given for interaction effects indicate significant interactions in models including all main and interaction effects.

6.3.8 Summary of nature of mode effects

To summarise, we find no evidence that the showcards lead to differences in satisficing behaviour (no support for H1 and H2) or that there is more satisficing in telephone interviews than in face-to-face interviews with no showcards (no support for H3 and H4). On the contrary, we find the opposite effect: more satisficing with face-to-face interviews carried out without showcards. We also find no evidence that differences in satisficing behaviour across modes differ by cognitive ability (no support for H5). We do however find that responses that might be considered to be socially desirable are more likely to be given over the telephone than in interviews conducted face-to-face without showcards (no support for H6, H7 supported) and we also find that there are differences in the likelihood of choosing extreme response categories (telephone respondents selected fewer last-category responses), although these do not appear to be explained by the use of showcards. Rather, less extremeness appears to be a further manifestation of social desirability bias in these data.

For many items the differences between face-to-face interviews with showcards and telephone interviews are not found when broken down into differences due to showcards or

due to interviewer presence. This may suggest that the effects of showcards and interviewer presence are multiplicative rather than independent.

6.4 Respondents' experience of the survey interviews

A further indicator of the impact of mode of data collection on survey quality is provided by measures of the respondents' experience of the interviews by mode. Table 20 shows the mean duration of the interview in each of the treatment groups. Telephone interviews were significantly shorter than face-to-face interviews. The use of showcards did not significantly affect the total duration of the face-to-face interviews. Respondents' perception of the length of the interview was measured by asking whether they would have been willing to continue being interviewed for longer. Table 21 shows that telephone respondents were significantly more likely than face-to-face respondents to be willing to continue the interview for longer and less likely to say they did not wish to continue. Education was not found to be a significant predictor of willingness to continue the interview for longer.

Table 20: Mean interview duration (minutes)

Telephone	F2f showcard	F2f no showcard	Telephone vs. f2f showcard	F2f no showcard vs. f2f showcard	Telephone vs. f2f no showcard
15.34	17.43	17.64	***	ns	***

Notes: Asterisks indicate P-values of the mode coefficient from regressions of interview duration on mode and socio-demographics.

Table 21: Respondents' perception of interview length

Willing to continue interview…	F2f showcard	F2f no showcard	Telephone	Telephone vs. f2f showcard	F2f no showcard vs. f2f showcard	Telephone vs. f2f no showcard
Much longer	11.65	15.25	20.06	omitted	omitted	omitted
A bit longer	45.44	42.86	48.49	*	ns	ns
not at all	39.42	38.22	30.67	***	ns	**
DK/refused	3.50	3.67	0.78	-	-	-

Notes: P-values from multinomial models of whether willing to continue interview on mode and socio-demographics, where 'don't know/refused' category excluded. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, ns=not significant.

Respondents were asked what their preferred mode of data collection would be if they were asked to participate in an hour-long interview in their own home. The results are shown

in table 22. Mode preferences tend to be biased by the mode in which the question is administered and this pattern is also replicated in these data (see Groves, 1979). Face-to-face respondents were significantly more likely to express a preference for a face-to-face interview, whereas telephone respondents were more likely to express a preference for alternative modes.

Finally, respondents were asked whether they felt uneasy about answering questions on certain topics (table 23). This provides a crude measure of the sensitivity of the topics (and thereby, of their susceptibility to social desirability bias), but also of how at ease respondents felt being interviewed in different modes. Income and voting were the topics respondents felt most uneasy about, followed by immigration and politics. Compared with telephone respondents, respondents interviewed face-to-face were more likely to say they felt uneasy about answering questions on certain topics, suggesting that overall the in-person interviews were sometimes uncomfortable for respondents. However, this discomfort did not translate into increased social desirability bias among face-to-face respondents. Rather, respondents who found the questions to be sensitive were more likely to have responded truthfully. Note that these items themselves may have also have been affected by social desirability bias.

Table 22: Preferred mode for hypothetical one hour survey in respondent's home

	F2f showcard	F2f no showcard	Telephone	Telephone vs. f2f showcard	F2f no showcard vs. f2f showcard	Telephone vs. f2f no showcard
F2f	24.27	25.87	8.18	omitted	omitted	omitted
Telephone	14.76	15.44	32.52	***	ns	***
Paper self-completion	22.14	17.95	26.39	***	ns	***
Web	12.43	13.71	21.13	***	ns	***
Other	3.50	3.67	5.55	ns	ns	ns
Refused	17.67	16.80	4.09			
Don't know	5.24	6.56	2.14			

Notes: Raw distribution of mode preference. P-values from multinomial models of preferred mode on mode and socio-demographics, where 'other, refused and don't know' categories were combined. * $p<0.05$, ** $p<0.01$, *** $p<0.001$, ns=not significant.

Table 23: Percentage reporting unease about answering questions on certain topics

	F2f showcard	F2f no showcard	Telephone	Telephone vs. f2f showcard	F2f no SCvs. F2fSC	Telephone vs. f2f no showcard
Income	22.72	31.47	17.04	*	**	***
Immigration	10.87	7.14	7.40	ns	*	ns
Politics	7.96	9.46	5.84	ns	ns	ns
Voting	15.34	15.06	8.67	***	ns	***
Religion	6.60	6.18	2.63	**	ns	**
Gender roles	3.69	3.09	1.36	*	ns	*

Notes: Raw percentage uneasy answering questions, where 'yes', 'don't know' and 'not answered' were coded as 1 and 'no' coded as 0. Asterisks indicate P-values of the mode coefficient from logit regressions of unease on mode and socio-demographics. * p<0.05, ** p<0.01, *** p<0.001, ns=not significant.

Table 24: Interviewer observations

	F2f showcard	F2f no showcard	Telephone	Telephone vs. f2f showcard	F2f no showcard vs. f2f showcard	Telephone vs. f2f no showcard
Asked for clarification?						
Never	46.41	46.72	60.27	***	ns	***
Almost never	29.71	26.25	20.84			
Now and then	19.81	24.13	16.65			
Often	3.88	2.12	1.85			
Very often	0.19	0.77	0.39			
Reluctant to answer?						
Never	52.05	43.99	61.31	***	ns	***
Almost never	23.20	21.90	17.15			
Now and then	17.93	26.36	17.74			
Often	5.07	5.04	2.34			
Very often	1.75	2.71	1.46			
Answered to best ability?						
Never	0.39	0.39	0.49			
Almost never	0.39	0.59	0.20			
Now and then	5.08	5.27	3.50			
Often	34.38	33.20	17.48			
Very often	59.77	60.55	78.32	***	ns	***
Understood questions?						
Never	0.19	0.58	0.49			
Almost never	0.58	0.39	0.49			
Now and then	4.87	5.43	3.91			
Often	31.58	31.78	18.57			
Very often	62.77	61.82	76.54	***	ns	***
Influenced by other person?						
Yes	5.44	7.34	2.14	**	ns	***
Distracted?						
Yes/maybe/dk	8.74	11.58	15.09	***	ns	*

Notes: Asterisks indicate P-values from logit models of 'never' or 'very often' versus all other categories on mode and socio-demographics. * p<0.05, ** p<0.01, *** p<0.001, ns=not significant.

Interviewers were asked six questions about the extent to which respondents were engaged in the interview, how well they had understood the questions and whether there were any distractions present during the interview. Results are shown in table 24. Significant differences between ratings from telephone and face-to-face interviewers were observed for all six items. According to the interviewers assessments, more telephone than face-to-face respondents 'never' or 'almost never' requested clarification on questions or showed reluctance to answer questions; and 'often' and 'very often' tried to answer the questions to the best of their ability and understood the questions. Unsurprisingly, it was also less common for another person(s) present to influence the responses given in telephone interviews. However, telephone respondents were more likely to appear distracted during the interview. Note that telephone interviewers do not have the same information (such as nonverbal cues) available to them during the course of the interview as face-to-face interviewers, so their observations are restricted to verbal and auditory signals. There were no differences observed between the two face-to-face groups.

7 Discussion and conclusion

7.1 Summary of analysis and results

Data from the phase 2 experiment were analysed for evidence of mode effects in terms of differences between the three treatment groups attributable to characteristics of the data collection procedures used in each. Of the three groups, two were interviewed in person, and one was interviewed by telephone. The two face-to-face groups differed by the use of showcards in the standard ESS 'control' condition and the use of the questionnaire adapted for telephone in the no-showcard condition. This allowed us to compare telephone interviewing with the standard ESS method, whilst ensuring strict comparability between the groups.

The principal differences between face-to-face and telephone interviewing concern the 'channels of communication' available in each mode and the 'level of intimacy' (Groves and Kahn, 1979) between the interviewer and respondent. For example, face-to-face interaction allows the use of visual cues, including nonverbal communication and the use of showcards. Both are argued to facilitate the interview process (e.g. see Holbrook et al. 2003) by helping to engage and motivate respondents and reduce the cognitive burden of the survey task. The physical presence of the interviewer and the availability of nonverbal cues also

play a role in the build-up of rapport between interviewers and respondents in face-to-face interviews, making in-person interaction more intimate than that conducted by telephone.

Based on these key differences between the modes, we anticipated variation in the quality of data from each of the three treatment groups. To assess the impact of mode on data quality, we first examined the magnitude of mode effects observed in the data. This was done by comparing the similarity of response distributions across 33 items in the questionnaire. Two methods were used: 1) a regression approach to isolate the effect of mode on mean scores while controlling for differences in the demographic compositions of the samples; 2) a proportional odds modelling technique to compare the distribution of responses across each of the response categories for items measured ordinally. This allowed us to identify those items most likely to be affected by a switch to telephone interviewing. Of the 33 items tested, just 13 were affected by mode of interviewing. In 8 cases, we observed significant differences between face-to-face and telephone interviews, but no differences between the two face-to-face groups, implying that the presence of the interviewer was a more important factor than the sensory channel (aural vs. visual) in determining the likelihood of mode effects. Just two items exhibited mode effects that could be attributed to the sensory channel. The results of the proportional odds models broadly mirrored these findings.

In all but one case (household income), the observed mode effects were comparatively small (no greater than one standard error) such that they would be unlikely to affect the conclusions of analysts using mixed mode data. To test this further, we examined what impact the observed mode effects had on summed attitude scores (from scale items) and on the relationship between variables. Mode effects for summed scores mirrored those found at the item level, while the relationships between variables remained unaffected.

At the next stage of the analysis, we examined the nature of the observed mode differences and tested hypotheses about the likely causes of the mode effects. Mode can impact on data quality in a variety of ways: for example, it can affect the completeness of data (including patterns of item non-response and the length of responses to open questions) and it can bias the substantive responses given. Two types of bias are likely to be of especial concern if the ESS were to switch to telephone interviewing. Firstly, because of the greater cognitive demands of audio-only communication, together with the typically faster pace with which telephone interviews are conducted, telephone respondents (particularly those who are less motivated or able) are likely to 'shortcut' cognitive steps involved in the response process and adopt satisficing strategies to answer the questionnaire, such as using the same

48

scale point to rate a number of different objects (so-called *non-differentiation*); always agreeing with items (*acquiescence*); or showing a preference for response options near the start of a list when they are presented visually (e.g. on a showcard) and towards the end of a list presented aurally (so-called *primacy* and *recency / response order effects*).

Secondly, because of the reduced level of intimacy between the interviewer and respondent in telephone interviews, respondents may feel less inclined to answer truthfully where questions are of a sensitive nature. Instead, the respondent may select a response that is more socially desirable in order to portray themselves in a more favourable light to the interviewer. We tested a range of hypotheses derived from these observations from previous research in the field. In H1, we tested whether the use of showcards helped to simplify the response task by looking at evidence for respondent satisficing across the two face-to-face groups. In H2, we tested in particular whether showcards led to primacy effects and whether questions without showcards are more susceptible to recency effects. Overall, we found no significant differences between the two face-to-face groups in terms of non-differentiation, acquiescence and response order effects, suggesting that showcards have little effect on the quality of the data collected. No support was found for either H1 or H2.

In H3 and H4, we tested whether the presence of the interviewer helps to reduce task difficulty and to increase respondent motivation (respectively), once again by looking at evidence for satisficing, this time between the telephone and face-to-face groups (recall that only the no-showcard and telephone group are strictly comparable). We found no evidence of increased non-differentiation among telephone respondents. However, there were some differences between the telephone and showcard groups, suggesting that while the likelihood of non-differentiation may not be influenced separately by the presence of the interviewer or the use of showcards, it may result from the interaction between the two. There was no evidence of increased acquiescence among telephone respondents (in fact face-to-face respondents interviewed without showcards were most likely to acquiesce), nor did we find any evidence of response order effects. No support was found for either H3 or H4.

In H5, we tested the hypothesis that the impact of mode on the propensity to shortcut would be greatest for respondents with low cognitive ability. This was also not supported by the evidence.

H6 and H7 addressed the likelihood of socially desirable reporting in face-to-face and telephone interviews. H6 assumes that the increased anonymity of telephone interviews and the social distance between the interviewer and respondent will make respondents *more* likely to report socially undesirable attitudes and behaviours over the telephone. In H7, the

assumption is that the enhanced rapport in face-to-face interviews will encourage respondents to be more honest in reporting their socially undesirable attitudes and behaviours. We found evidence in support of H7 – i.e. more social desirability bias among telephone respondents.

At the final stage of our analysis, we assessed respondents' experiences of the interview in each mode and their preference for different modes. Interviews conducted by telephone were significantly shorter than those conducted in person and telephone respondents were more likely to be willing to continue the interview for longer. Respondents were also asked whether they felt uneasy about answering questions on certain topics. Face-to-face respondents were more likely to report that they had felt uneasy, although as noted, this did not lead them to bias (in the direction of social desirability) their responses to questions on topics they found sensitive. Finally, based on observations recorded by the interviewer, telephone respondents were less likely to ask for clarification and less likely to be reluctant to answer; they were also rated as more likely to answer questions to the best of their ability and more likely to have understood the questions. However, it is clear that interviewers' observations were themselves influenced by the mode in which the interview was conducted.

7.2   Implications of findings for the ESS

7.2.1   Differential non-response by mode

One of the least surprising, yet perhaps most concerning findings of this research is that people who agree to participate in telephone interviews differ systematically from those who agree to be interviewed face-to-face. The analytical methods used here were necessary to control for significant differences in the samples with respect to level of education, proportion of manual workers and sex. Once these differences were controlled for, however, the relative effect of mode was minor (affecting only around one third of the items in the questionnaire) and it did not appear to have a serious impact on the relationship between variables, suggesting that data users could be confident in the results of their statistical analyses of mixed telephone and face-to-face data. Nevertheless, the onus would be on the data analyst to use appropriate techniques to take account of differential non-response between samples. This poses an important challenge for the ESS when considering the move to multimode data collection. Understanding the causes of differential non-response (coverage errors, selection biases, method of contact, public preferences for different modes, etc.) will be an important part of this work. It will also be necessary to consider how to

accommodate this 'by product' of mixing modes – for example, developing appropriate weighting procedures for non-response, understanding the potential biasing effects on the data and so on; but also to consider how to take advantage of it – mixing modes may make it possible to achieve a more representative sample overall.

7.2.2 Social desirability bias

A growing body of empirical evidence lends support to the theory that failure to establish rapport in telephone interviews (and the fewer opportunities for interviewers to convince respondents of the legitimacy of the survey) leads respondents to answer questions less honestly than when they are interviewed in person. The results of our research provide compelling further evidence of this effect. Attitudinal measures on a range of topics that we believed might be governed by powerful social norms consistently yielded more socially desirable responses from telephone respondents than from face-to-face respondents, making this a particularly concerning source of mode effects if the ESS were to permit a switch to telephone interviewing. Understanding more about how this type of bias operates, therefore, represents an important area for further research.

In particular, not much is known about the cognitive mechanisms underlying social desirability bias. A number of theories exist, the most compelling of which suggests that social desirability bias results from respondents editing their true response to survey questions in light of impression management concerns (see Tourangeau, Rips and Rasinski, 2000, for an overview). The logical extension of this is that response times to sensitive questions will be longer than those for more neutral questions because the respondent must engage in greater cognitive effort to assess their true response in relation to the social desirability connotations of a question and modify their answer accordingly to portray themselves to the interviewer more favourably. Indeed, there is some evidence to support this (e.g. Holtgraves, 2004). Yet this explanation does not tally with the finding that telephone interviews are generally conducted at a faster pace than face-to-face interviews and that they carry a greater cognitive burden for respondents, giving respondents less opportunity to think carefully about their answers.

An alternative explanation might be that respondents select the most socially desirable response because it is the easiest, most accessible or salient response available to them (without having to engage in 'deep' processing) – a theory that has been used to explain

acquiescent response bias (Knowles and Condon 1999). If this is the case, response latencies would be correspondingly shorter, suggesting that the apparent social desirability bias is in fact another form of shortcutting in its own right. Understanding these cognitive processes better will be fundamental to developing methods of mitigating the bias in telephone interviews.

A further implication of the findings relating to social desirability bias concerns the extent to which the results would be replicated cross-nationally. Comparatively little is known about cultural differences in the manifestation of social desirability bias, such as variations in the connotations of particular topics, the social norms governing certain types of behaviour and the importance of impression management strategies in different countries. Again, this constitutes an important avenue for future research.

7.2.3   Respondent satisficing

The fact that our questionnaire was relatively short makes it perhaps unsurprising that we found little evidence of mode effects on satisficing. Holbrook et al. (2003) found consistent evidence of an increase in satisficing among telephone respondents – and particularly among those with low education. However, their study was specifically concerned with surveys using long questionnaires. Our questionnaire was not only quite short, but also quite varied, and the topics changed frequently, so it was perhaps more stimulating for respondents compared to some longer surveys handling fewer subtopics (or with more questions on each subtopic). Given that response rates were comparatively low, we might also assume that those sample members who did take part would be more willing to expend the necessary effort to respond to the questions thoughtfully than our non-respondents would have. So we should be cautious about interpreting these findings as evidence that satisficing is not a problem on the ESS, nor that it would not be in a telephone version of the survey. On the contrary, given the fact that the ESS is a long survey (lasting on average one hour to administer in person, though considerably longer in some countries), with relatively long modules of questions on the same topic, as well as the fact that samples are likely to contain a higher proportion of reluctant respondents we should be particularly wary of satisficing effects in ESS data. On the surface it is reassuring that our data were relatively free from this bias, however, we acknowledge the risk of drawing inferences to the ESS as a whole, given that a reduced version of the questionnaire does not provide an ideal method of testing the theory.

Exploring the impact of interview length on satisficing in telephone surveys (and indeed, the feasibility of carrying out long survey interviews by telephone) will be fundamental to the decision about whether a switch to telephone mode will be permissible on the ESS.

7.2.4   Questions susceptible to mode effects

One of the aims of this research was to be able to identify which questions or types of questions in the ESS would be most sensitive to mode. We selected items from the core questionnaire, which – based on the findings of previous research in the field – we assumed would be especially susceptible to mode effects, on the grounds that if no effects were observed in our experiment for these items, we could be reasonably confident about the mode sensitivity of the excluded questions. Some items were selected because they were deemed to have widely shared social desirability connotations (although we did not test this empirically in the countries where the research was carried out), others because – together with other items – they formed part of a scale that might be likely to induce forms of respondent satisficing under certain conditions. Other questions were chosen according to their level of complexity; others because there were open questions; others because they were likely to be problematic due to the number of response categories exhibited on the showcard in the standard ESS interview. This latter type of questions posed a particular challenge in terms of how best they should be modified to make them suitable for telephone administration. In each case, we tried to keep changes to the form of the questions and the response categories to a minimum. However, for a small set of questions, the adaptations were more substantial. For example, three items in the question asked about quantities: how much time respondents spend watching TV and TV news and how much is their household income. For all three, the telephone and no-showcard solution was to ask respondents outright for an estimate of the quantity involved.

Overall, we can be reassured that only a third (13) of the questions were affected by mode effects and that the effects we observed were relatively small. We can be further assured by the fact that of these, only a small number of these effects appeared to result from differences in the nature of question stimulus – either because of the use of the visual stimulus of showcards, or because of substantial modifications to the aural versions of the questions. Of these, one was the income question (which exhibited significant differences between all three treatment groups) and one was time spent watching television (where a

difference was observed between the two face-to-face groups, suggesting it was attributable to the change in question form and/or the use of showcards). Setting aside the complex issues concerning the measurement of income per se, as well as how to measure it in different modes, these findings highlight the potential danger of adapting questions for which showcards are used to aural modes. Future work could usefully be focused on testing variations in question form, in order to achieve stimulus equivalence across different modes of data collection. This will be particularly important in the development of an appropriate measure of household income because of the difficulty of disentangling the interactive effects of interviewer presence and the use of showcards (which in the case of income measurement, are also designed to promote confidentiality).

The majority of the effects we observed, however, were consistent with social desirability bias. Questions about topics for which there are shared assumptions about what views it is socially acceptable and unacceptable to voice publicly, seem especially likely to be susceptible to the bias, so any decision to move to telephone interviewing should be conscious of the potential damaging effect this could have on continuity and cross-cultural comparability. However, as noted, a clear limitation of the study was the absence of any empirical test of the social desirability connotations of the questions in the countries concerned (though our measure of respondent unease in relation to certain topics was partly intended to address this issue), so a priority for the ESS should be to establish which questions and topics are likely to be most at risk of social desirability bias and the extent to which this varies cross-nationally.

7.3 Recommendations for mitigating mode effects

A second key aim of this research programme is to develop recommendations about how mode effects might be mitigated in a multimode survey. The rationale behind the design of our research was first to learn more about the causes of any observed effects in order to understand better how the effects might be minimised or prevented. Measurement errors can be attributed to the questionnaire, the interviewer and the respondent. Our design enabled us to distinguish between just two of these: mode effects that could be attributed to the nature of the question stimulus (i.e. the use of showcards, or differences in question form between the visual and aural modes) and effects that were attributable to other differences in the mode – most notably, the presence or absence of the interviewer, but also other characteristics of the

mode and 'bundle' of methods associated with collecting data in that mode (Groves, 1979) that we did not explicitly control for in the design of this experiment.

In general, we were successful at developing equivalent aural versions of questions normally administered with the help of showcards on the ESS, by keeping modifications to a minimum. However, as noted problems arose with respect to the measurement of income and the amount of time spent watching television. For this reason, further experimentation with alternative versions of these and similar questions will be necessary in order to find more equivalent forms for use in mixed mode surveys.

In order to minimise the likelihood of social desirability bias in telephone data, it is essential that we understand better the causes and mechanisms underlying the bias. Traditionally, the solution to the problem has been to offer the respondent more privacy in the response process (e.g. self-completion methods) to ensure the confidentiality of their responses. Paradoxically, the social distance between the respondent and interviewer in a telephone survey ought to encourage a greater sense of anonymity than a face-to-face interview does, yet we find more social desirability in the former. It is possible that the effect might be minimised if the respondent could be reassured of the legitimacy of the survey (and about confidentiality issues) – perhaps through an advance letter or some other method (such as a specially scripted introduction from the interviewer). Research (including a review of existing related studies) will be necessary to identify which method(s) have the most positive impact on data quality. Similarly, understanding better the cognitive processes involved will further inform decisions about how to reduce the likelihood and impact of the effect. Readers should refer to Holbrook, Green and Krosnick (2003) for a discussion of methods of minimising the likelihood of satisficing in telephone surveys.

Finally, how might we mitigate mode effects relating to coverage, selection and non-response bias? This perhaps represents the area most in need of further study – particularly in a comparative context where few attempts have been made to pool what is currently known about cross-national variations in the penetration of modes, the availability of suitable frames, public preferences for different modes, survey recruitment methods and so on[6]. Only by building up our understanding of survey practice in different countries can we begin to recommend best practice for carrying out mixed mode surveys and make decisions about the

---

[6] An exception is provided by a study of survey practices in European countries participating in the International Adult Literacy Survey ONS. 2000. Measuring adult literacy: the International Adult Literacy Survey in the European Context. London: Office for National Statistics.

feasibility of fielding a cross-national survey in different modes - either separately or in combination.

8 References

AAPOR. 2006. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 4th ed. Lenexa, Kansas: AAPOR.

Aneshensel, C.S., R.R. Frerichs, V.A. Clark, and P.A. Yokopenic. 1982. Measuring Depression in the Community: A Comparison of Telephone and Personal Interviews. *Public Opinion Quarterly* 46 (1):110-121.

Aquilino, W.S. 1994. Interview mode effects in surveys of drug and alcohol use: A field experiment. *Public Opinion Quarterly* 58 (2):210-240.

Billiet, J., and H. Welkenhusen-Gybels. 2004. Assessing cross-national construct equivalence on the ESS: The case of six immigration items. Paper read at Sixth International conference on Social Science Methodology, at Amsterdam.

Bishop, G., H.J. Hippler, N. Schwarz, and F. Strack. 1988. A comparison of response effects in self-administered and telephone surveys. In *Telephone survey methodology*, edited by R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II and J. Wakesberg. New York: Wiley.

Braun, M. 2003. Errors in comparative survey research: An overview. In *Cross-cultural survey methods*, edited by J. A. Harkness, F. J. R. Van der Vijver and P. P. Mohler. Hoboken, NJ: John Wiley.

Cassel, Carol. A. . 2004. Voting records and validated voting studies. *Public Opinion Quarterly* Vol. 68 (No. 1):102-108.

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

de Leeuw, E. 2005. To mix or not to mix? Data collection modes in surveys. *Journal of Official Statistics* 21 (2):1-23.

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

DeMaio, T.J. 1984. Social desirability and survey measurement: A review. In *Surveying subjective phenomena: Volume 2*, edited by C. F. Turner and E. Martin. New York: Russell Sage.

Dijkstra, W., and Y.P. Ongena. 2002. Question-answer sequences in survey-interviews. Paper read at International Conference on Questionnaire Development, Evaluation and Testing Methods, at Charleston, SC.

Dillman, D.A. 2000. *Mail and internet surveys: The tailored design method*. 2nd ed. New York: John Wiley Co.

Dillman, D.A., T.L. Brown, J.E. Carlson, E.H. Carpenter, F.O. Lorenz, R. Mason, J. Saltiel, and R.L. Sangster. 1995. Effects of category order on answers in mail and telephone surveys. *Rural Sociology* 60 (4):674-687.

Dillman, D.A., J.R. Clark, and K.K. West. 1995. Influence of an invitation to answer by telepohone on response to census questionnaires. *Public Opinion Quarterly* 58:557-568.

Dillman, D.A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, and J. Berck. 2000. Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the internet. Paper read at Annual Conference of the American Association for Public Opinion Research, at Montreal, Quebec, Canada.

Dillman, D.A., R.L. Sangster, J. Tarnai, and T. Rockwood. 1996. Understanding differences in people's answers to telephone and mail surveys. In *Advances in survey research*, edited by M. T. Braverman and J. K. Slater. San Francisco: Jossey-Bass.

Dillman, D.A., and J. Tarnai. 1991. Mode effects of cognitively-designed recall questions: A comparison of answers to telephone and mail surveys. In *Measurement errors in surveys*, edited by P. P. Beimer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz and S. Sudman. New York: Wiley.

Duffy, B. 2003. Response order effects - How do people read? *International Journal of Market Research* 45 (4):457-466.

Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. An experimental comparison of web and telephone surveys. *Public Opinion Quarterly* 69 (3):370-392.

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

Groves, R.M. 1989. *Survey errors and survey costs*. New York: Wiley.

Groves, R.M. 1990. Theories and methods of telephone surveys. *Annual Review of Sociology* 16:221-240.

Groves, Robert.M., and Robert.L. Kahn. 1979. *Surveys by telephone: A national comparison with personal interviews*. New York, NY: Academic Press.

Hadaway, C.K., P.L. Marler, and M. Chaves. 1993. What the polls don't show: A closer look at U.S. church attendance. *American Sociological Review* Vol. 58 (No. 6):741-752.

Hochstim, J.R. 1967. A critical comparison of three strategies of collecting data from households. *Journal of the American Statistical Association* 62:976-989.

Holbrook, A. L., M. C. Green, and J. A. Krosnick. 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.

Holtgraves, T. 2004. Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin* 30 (2):161-172.

Japec, L. 1995. *Issues in mixed-mode survey design.* Stockholm: Statistics Sweden.

Jordan, L.A., A.C. Marcus, and L.G. Reeder. 1980. Response styles in telephone and household interviewing: A field experiment. *Public Opinion Quarterly* 44:210-222.

Jowell, R. 1998. How comparative is comparative research? *Amercian Behavioural Scientist* 42 (2):168-177.

Kalton, G., D. Kasprzyk, and D. B. McMillen. 1989. Nonsampling errors in panel surveys. In *Panel Surveys*, edited by D. Kasprzyk. New York: John Wiley.

Karp, J.A., and D. Brockington. 2005. Social desirability and response validity: A comparative analysis of overreporting voter turnout in five countries. *The Journal of Politics* 67 (3):825-840.

Knowles, E.S., and C.A. Condon. 1999. Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Applied Psychology* 77 (2):379-386.

Körmendi, Eszter. 1988. The quality of income information in telephone and face to face surveys. In *Telephone survey methodology*, edited by R. M. Groves, P. P. Biemer, L. E. Lyberg, J. T. Massey, W. L. Nicholls II and J. Waksberg. New York: John Wiley and Sons.

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. 1999. Survey research. *Annual Review of Psychology* 50:537-567.

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

Krosnick, J. A., S. Narayan, and W.R. Smith. 1996. Satisficing in surveys: Initial evidence. In *Advances in Survey Research*, edited by M. T. Braverman and J. K. Slater. San Francisco: Jossey-Bass Publishers.

Krysan, M., H. Schuman, L.J. Scott, and P. Beatty. 1994. Response rates and response content in mail versus face-to-face surveys. *Public Opinion Quarterly* 58 (3):381-399.

Locander, W.B., and J.P. Burton. 1976. The effect of question form on gathering income data by telephone. *Journal of Marketing Research* Vol. 13 (No. 2, May):189-192.

Long, J. S. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Lyberg, L.E., and D. Kasprzyk. 1991. Data collection methods and measurement error: An overview. In *Measurement errors in surveys*, edited by P. P. Biemer, R. M. Groves and L. E. Lyberg. New York: John Wiley.

Lynn, P., S. Gabler, S. Häder, and S. Laaksonen. 2004. Methods for Achieving Equivalence of Samples in Cross-National Surveys. *ISER Working Paper* 2004-09. Colchester: University of Essex.

Miller, Peter.V. 1984. Alternative question forms for attitude scale questions in telephone interviews. *Public Opinion Quarterly* 48:766-778.

Mooney, G., L. Giesbrecht, and C. Shettle. 1993. To pay or not to pay: That is the question. Paper read at Paper presented at the meeting of the Amercian Association for Public Opinion Research, at St. Petersburg, FL.

Moore, D.L. 1998. *1998 Survey of grass seed growers in Washington state (Data Report No. 98-06)*. Pullman, WA: Washington Statey University, Social and Economic Sciences Research Center.

Narayan, Sowmya, and Jon A. Krosnick. 1996. Education moderates some response effects in attitude measurement. *Public Opinion Quarterly* Vol. 60:58-88.

O'Connell, A.A. 2006. *Logistic regression models for ordinal response variables*. Thousand Oaks, CA: Sage.

ONS. 2000. Measuring adult literacy: the International Adult Literacy Survey in the European Context. London: Office for National Statistics.

Peytcheva, Emilia A., Robert Manchin, Robert Tortora, and Robert.M. Groves. 2004. Comparing face to face, telephone, paper self-administered and web survey measurement. Paper read at American Association for Public Opinion Research 59th Annual Conference, May 13th-16th, at Phoenix, Arizona.

Rogers, T.F. 1976. Interviews by telephone and in person: quality of responses and field performance. *Public Opinion Quarterly* 40:51-65.

Saris, W.E. 1997. Adjustment for differences between face to face and telephone interviews. In *Eurobarometer - Measurement instruments for opinions in Europe*, edited by W. E.

Saris and M. Kaase. Mannheim: Zentrum fuer Umfragen, Methoden und Analysen (ZUMA).

Schaefer, D., and D.A. Dillman. 1998. Development of a standard e-mail methodology: Results of an experiment. *Public Opinion Quarterly* 62:378-397.

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

Schwarz, N., H.J. Hippler, and E. Noelle-Neumann. 1992. A cognitive model of response-order effects in survey measurement. In *Context effects in social and psychological research*, edited by N. Schwarz and S. Sudman. New York: Springer-Verlag.

Silver, Brian, D., Barbara Anderson, A., and Paul Abramson, R. 1986. Who overreports voting? *American Political Science Review* Vol. 80 (No. 2 (June)):613-624.

Sykes, Wendy, and Martin Collins. 1988. Effects of mode of interview: Experiments in the UK. In *Telephone Survey Methodology*, edited by R. M. Groves, P. P. Biemer, L. Lyberg, E., J. T. Massey, W. L. Nicholls II and J. Waksberg. New York: Wiley and Sons, Inc.

Sykes, Wendy, and G. Hoinville. 1985. *Telephone interviewing on a survey of social attitudes: A comparison with face-to-face procedures*. London: Social and Community Planning Research.

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

Tourangeau, R., and T.W. Smith. 1996. Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly* 60 (2):275-304.

Weeks, M.F. 1992. Computer-assisted survey information collection: A review of CASIC methods and their implications for survey operations. *Journal of Official Statistics* 8 (4):445-465.

Williams, R. 2006. Generalized ordered logit/ Partial proportional odds models for ordinal dependent variables. *The Stata Journal* 6 (1):58-82.

9 Appendix

9.1 Questionnaires

ESS-Gallup mixed mode methodology experiment – Phase 2

Face-to-face questionnaire (Group A: F2F + Showcards)

1. **CARD 1** On an average weekday, how much time, in total, do you spend watching television?

No time at all	00	**GO TO 3**
Less than ½ hour	01	
½ hour to 1 hour	02	
More than 1 hour, up to 1½ hours	03	
More than 1½ hours, up to 2 hours	04	**ASK 2**
More than 2 hours, up to 2½ hours	05	
More than 2½ hours, up to 3 hours	06	
More than 3 hours	07	
(Don't know)	88	

2. **STILL CARD 1** How much of this [response from 1] is spent watching **news** or programmes about **politics and current affairs**?

No time at all	00
Less than ½ hour	01
½ hour to 1 hour	02
More than 1 hour, up to 1½ hours	03
More than 1½ hours, up to 2 hours	04
More than 2 hours, up to 2½ hours	05
More than 2½ hours, up to 3 hours	06
More than 3 hours	07
(Don't know)	88

ASK ALL

3. **CARD 2** Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please use this scale from 0 to 10. 0 Means you can't be too careful and 10 means that most people can be trusted, and you can pick any number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
You can't be too careful										*Most people can be trusted*	*(Don't know)*
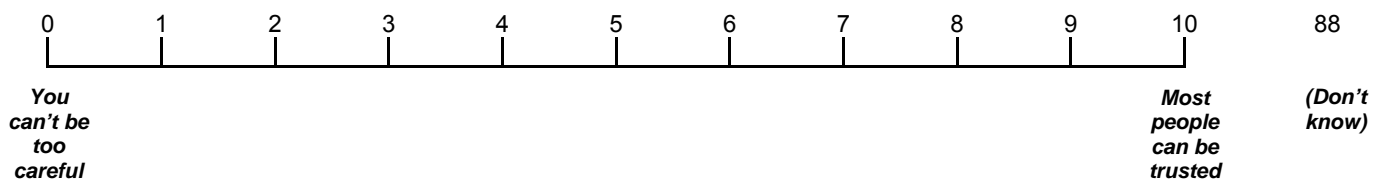
4. **CARD 3** All things considered, how satisfied are you with your life as a whole nowadays? Please use this scale from 0 to 10. 0 means extremely dissatisfied and 10 means extremely satisfied, and you can pick any number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88

Extremely dissatisfied   *Extremely satisfied*   *(Don't know)*
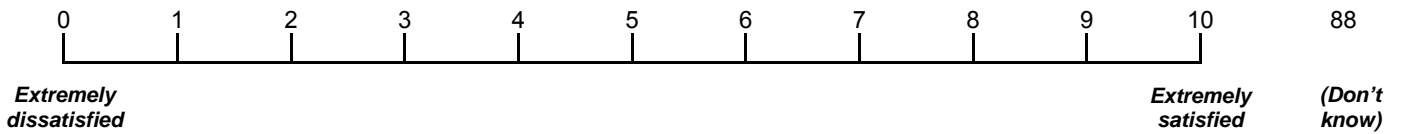
5. How interested would you say you are in politics – are you… **READ OUT**

very interested,	1
quite interested,	2
hardly interested,	3
or, not at all interested?	4
(Don't know)	8

6. **CARD 4** How often does politics seem so complicated that you can't really understand what is going on? Please use this card.

Never	1
Seldom	2
Occasionally	3
Regularly	4
Frequently	5
(Don't know)	8

7. **CARD 5** How easy or difficult do you find it to make your mind up about political issues? Please use this card.

Very difficult	1
Difficult	2
Neither difficult nor easy	3
Easy	4
Very easy	5
(Don't know)	8

8. **CARD 6** Using this card, please tell me on a score of 0-10 how much you <u>personally</u> trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly…**READ OUT**

		No trust at all										Complete trust	(Don't know)
a	…[country's] parliament?	00	01	02	03	04	05	06	07	08	09	10	88
b	…the legal system?	00	01	02	03	04	05	06	07	08	09	10	88
c	…the police?	00	01	02	03	04	05	06	07	08	09	10	88
d	…politicians?	00	01	02	03	04	05	06	07	08	09	10	88
e	… political parties?	00	01	02	03	04	05	06	07	08	09	10	88
f	…the European Parliament?	00	01	02	03	04	05	06	07	08	09	10	88
g	…the United Nations?	00	01	02	03	04	05	06	07	08	09	10	88

9. Some people don't vote nowadays for one reason or another. Did you vote in the last [country] national election in [month/year]?

Yes	1	**ASK 10**
No	2	
Not eligible to vote	3	**GO TO 11**
(Don't know)	8	

IF YES AT 9

10. Which party did you vote for in that election? [Country-specific (question and) codes]?

Party 1	01
Party 2	02
Party 3	03
Party 4	04
Party 5	05
Party 6	06
Other **(WRITE IN)** _____	07
(Refused)	77
(Don't know)	88

Now some questions about people from other countries coming to live in [country].

11. **CARD 7** Using this card, to what extent do you think [country] should allow people of the <u>same race or ethnic group</u> as most [country's] people to come and live here?

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

12. **STILL CARD 7** How about people of a <u>different</u> race or ethnic group from most [country] people? Still use this card.

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

13. **STILL CARD 7** How about people from <u>poorer countries outside Europe</u>? Use the same card.

Allow many to come and live here	1
Allow some	2
Allow a few	3
Allow none	4
(Don't know)	8

14. **CARD 8** Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries? Please use this card.

0	1	2	3	4	5	6	7	8	9	10	88
Bad for the economy										*Good for the economy*	*(Don't know)*
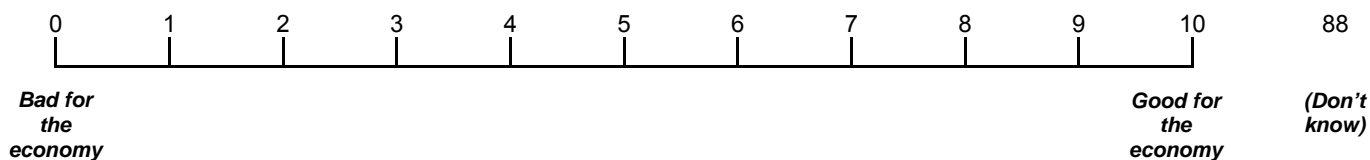
15. **CARD 9** And using this card, would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries?

0	1	2	3	4	5	6	7	8	9	10	88
Cultural life undermined										*Cultural life enriched*	*(Don't know)*
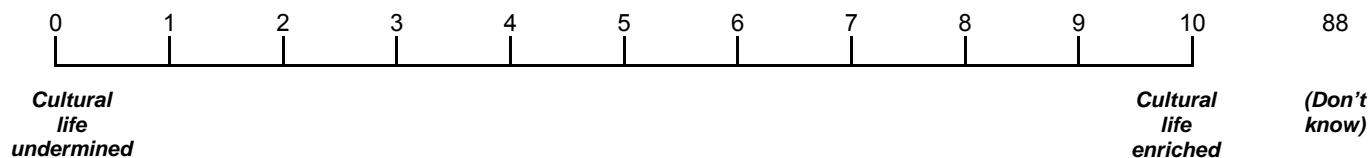
16. **CARD 10** Is [country] made a worse or a better place to live by people coming to live here from other countries? Please use this card.

0	1	2	3	4	5	6	7	8	9	10	88

Worse place to live   *Better place to live*   *(Don't know)*
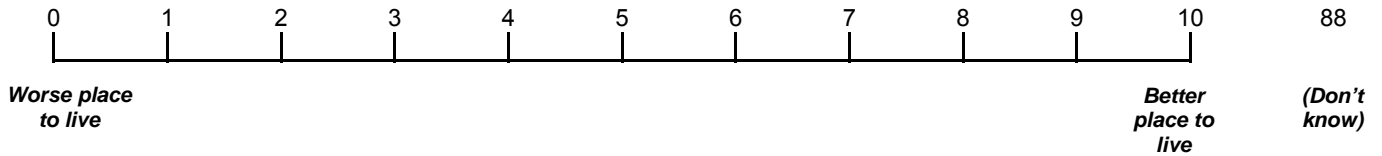
17. **CARD 11** I am now going to read out some statements about men and women and their place in the family. Using this card, please tell me how much you agree or disagree with the following statements. **READ OUT EACH STATEMENT AND CODE IN GRID**

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
a	A woman should be prepared to cut down on her paid work for the sake of her family	1	2	3	4	5	8
b	Men should take as much responsibility as women for the home and children	1	2	3	4	5	8
c	When jobs are scarce, men should have more right to a job than women	1	2	3	4	5	8
d	When there are children in the home, parents should stay together even if they don't get along	1	2	3	4	5	8

18. **STILL CARD 11** Still using this card, please say to what extent you agree or disagree with each of the following statements. **READ OUT EACH STATEMENT AND CODE IN GRID**

		Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly	(Don't know)
a	Gay men and lesbians should be free to live their own life as they wish	1	2	3	4	5	8
b	Whatever the circumstances, the law should always be obeyed	1	2	3	4	5	8

19. **CARD 12** Regardless of whether you belong to a particular religion, how religious would you say you are? Please use this card.

0	1	2	3	4	5	6	7	8	9	10	88

Not at all religious   *Very religious*   *(Don't know)*
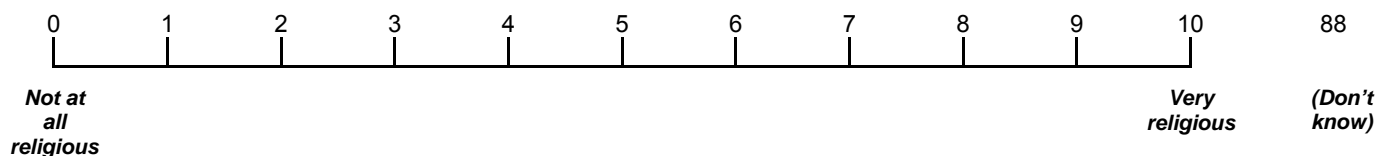
20. **CARD 13** Apart from special occasions such as weddings and funerals, about how often do you attend religious services nowadays?

Every day	1
More than once a week	2
Once a week	3
At least once a month	4
Only on special holy days	5
Less often	6
Never	7
(Don't know)	88

21. **CARD 14** Which phrase on this card best describes the area where you live?

A big city	1
The suburbs or outskirts of a big city	2
A town or small city	3
A country village	4
A farm or home in the countryside	5
(Don't know)	8

[Country-specific question and codes for coding into ESS Coding Frame]

22. **CARD 15** What is the highest level of education you have achieved? Please use this card.

	00
	01
COUNTRY SPECIFIC QUALIFICATIONS (SEE F6/ CARD 46)	02
	03
	04
	05
	06
(Don't know)	88

23. Can I just check, did you do any paid work (of an hour or more) in the last seven days?

Yes	1	**ASK 25**
No	2	**GO TO 24**

24. Have you ever had a paid job?

Yes	1	**ASK 25**
No	2	**GO TO 28**

ASK IF 1 AT 23 OR 1 AT 24

INTERVIEWER: If respondent currently in work (1 at 23), ask 25 about <u>current</u> job; if not in paid work but had a job in the past (1 at 24), ask 25 about <u>last</u> job.

25. What is/ was the name or title of your main job? **WRITE IN**

_____

26. In your main job, what kind of work do/did you do most of the time? **WRITE IN**

_____

_____

_____

27. **CARD 16** So, just to be sure, which of the descriptions on this card best describes the sort of work you do/did?

Senior professional	01
Other professional	02
Senior manager or administrator	03
Middle or junior manager or administrator	04
Other clerical	05
Technical and craft	06
Intermediate manual and service	07
Routine manual and service	08

ASK ALL

28. **CARD 17** People's income comes from lots of different sources, such as wages or salaries, pensions, social benefits, savings, investments and so on. Using this card, if you add up the income from **all** sources, which letter describes your household's total **net** income? If you don't know the exact figure, please give an estimate. Use the part of the card that you know best: weekly, monthly or annual income.

J	01
R	02
C	03
M	04
F	05
S	06
K	07
P	08
D	09
H	10
U	11
N	12

29. **INTERVIEWER CODE SEX OF RESPONDENT**

Male	1
Female	2

30. In which year were you born?

WRITE IN YEAR: | 1 | 9 | | |

31. Do you have access to the Internet at home or at work?

Yes, at home	1
Yes, at work	2
Yes, both at home and work	3
No	4

32. If you were asked to do a survey at home that would take about an hour, how would you choose to answer the questions? Would it be… **READ OUT**

…face-to-face interview,	1
telephone interview,	2
filling in a paper questionnaire,	3

69

or, some other way? **(WRITE IN)** _____   5

33. Some people feel a bit uneasy about some questions in surveys. I'll mention several types of question. Please tell me for each one whether or not you felt at all uneasy about answering them. **READ OUT EACH AND CODE IN GRID**

		Yes	No	(Don't Know)
a	Questions about your income	1	2	8
b	Questions about people from other countries coming to live in this country	1	2	8
c	Questions about politics in general	1	2	8
d	Questions about voting	1	2	8
e	Questions about religion	1	2	8
f	Questions about men and women's place in the home	1	2	8

34. Now we have finished the interview, I just want to ask you about the length of the interview. Would you have been willing to continue … **READ OUT**

… much longer,   1

a bit longer,   2

or not at all?   3

35. On an average weekday, how much time, in total, do you spend watching television?

WRITE IN   HOURS ☐☐   AND MINUTES ☐☐

36. How much of this [response from 1] is spent watching **news** or programmes about **politics and current affairs**?

WRITE IN   HOURS ☐☐   AND MINUTES ☐☐

ASK ALL

37. Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please use a scale from 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
You can't be too careful										*Most people can be trusted*	*(Don't know)*
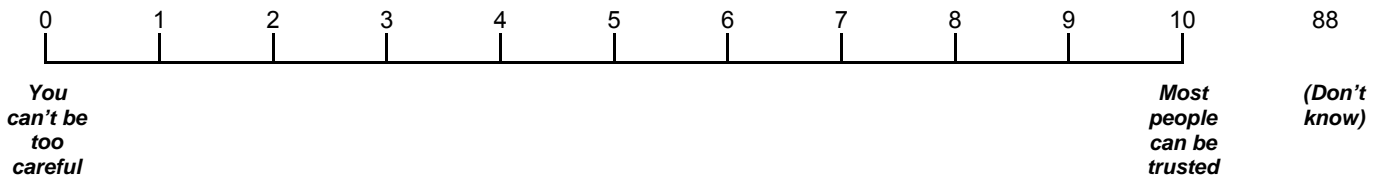
38. All things considered, how satisfied are you with your life as a whole nowadays? Please use a scale from 0 to 10, where 0 means extremely dissatisfied and 10 means extremely satisfied. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
Extremely dissatisfied										*Extremely satisfied*	*(Don't know)*
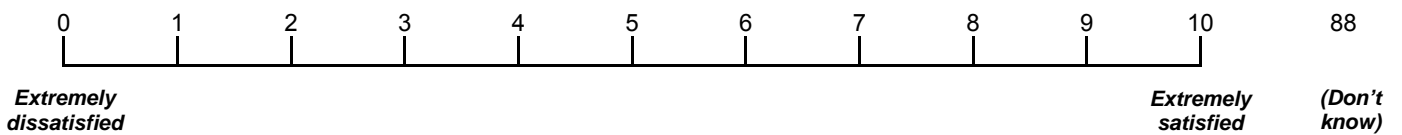
39. How interested would you say you are in politics – are you… **READ OUT**

very interested,	1
quite interested,	2
hardly interested,	3
or, not at all interested?	4
(Don't know)	8

40. How often does politics seem so complicated that you can't really understand what is going on? Would you say… **READ OUT**

never,	1
seldom,	2
occasionally,	3
regularly,	4
or, frequently	5
(Don't know)	8

41. How easy or difficult do you find it to make your mind up about political issues? Do you find it… **READ OUT**

very difficult	1
difficult	2
neither difficult nor easy	3
easy	4
or, very easy	5
(Don't know)	8

42. How much do you <u>personally</u> trust each of the following institutions? Please use a scale from 0 to10, where 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly…**READ OUT**

		No trust at all										*Complete trust*	*(Don't know)*
a	…[country's] parliament?	00	01	02	03	04	05	06	07	08	09	10	88
b	…the legal system?	00	01	02	03	04	05	06	07	08	09	10	88
c	…the police?	00	01	02	03	04	05	06	07	08	09	10	88
d	…politicians?	00	01	02	03	04	05	06	07	08	09	10	88
e	… political parties?	00	01	02	03	04	05	06	07	08	09	10	88
f	…the European Parliament?	00	01	02	03	04	05	06	07	08	09	10	88
g	…the United Nations?	00	01	02	03	04	05	06	07	08	09	10	88

43. Some people don't vote nowadays for one reason or another. Did you vote in the last [country] national election in [month/year]?

Yes	1	**ASK 10**
No	2	
Not eligible to vote	3	**GO TO 11**
(Don't know)	8	

IF YES AT 9

44. Which party did you vote for in that election? [Country-specific (question and) codes]?

Party 1	01
Party 2	02
Party 3	03
Party 4	04
Party 5	05
Party 6	06
Other **(WRITE IN)** _____	07
(Refused)	77
(Don't know)	88

Now some questions about people from other countries coming to live in [country].

45. To what extent do you think [country] should allow people of the <u>same race or ethnic group</u> as most [country's] people to come and live here? Do you think [country] should… **READ OUT**

…allow many to come and live here,	1
allow some,	2
allow a few,	3
or, allow none	4
(Don't know)	8

46. How about people of a <u>different</u> race or ethnic group from most [country] people? Do you think [country] should… **READ OUT**

…allow many to come and live here,	1
allow some,	2
allow a few,	3
or, allow none	4
(Don't know)	8

47. How about people from <u>poorer countries outside Europe</u>? Do you think [country] should… **READ OUT**

…allow many to come and live here,	1
allow some,	2
allow a few,	3
or, allow none	4
(Don't know)	8

48. Would you say it is generally bad or good for [country]'s economy that people come to live here from other countries? Please use a scale from 0 to 10, where 0 means it is bad for the economy and 10 means it is good for the economy. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
Bad for the economy										**Good for the economy**	*(Don't know)*
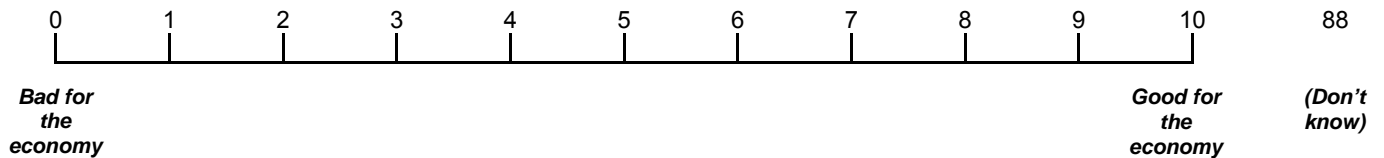
49. And would you say that [country]'s cultural life is generally undermined or enriched by people coming to live here from other countries? Use a scale from 0 to 10, where 0 means that [country]'s cultural life is undermined and 10 means [country]'s cultural life is enriched. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
Cultural life undermined										**Cultural life enriched**	*(Don't know)*
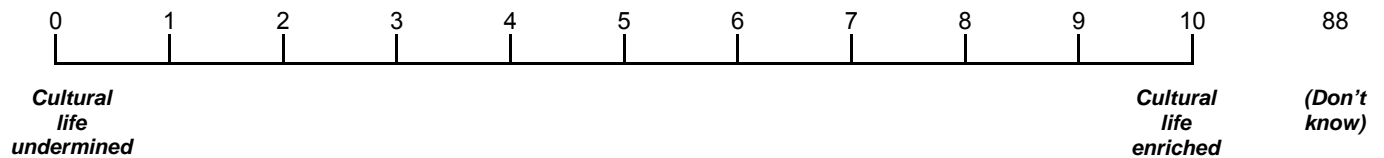
50. Is [country] made a worse or a better place to live by people coming to live here from other countries? Again, use a scale from 0 to 10, where 0 means [country] would be made a worse place to live and 10 means [country] would be a better place to live. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
Worse place to live										**Better place to live**	*(Don't know)*
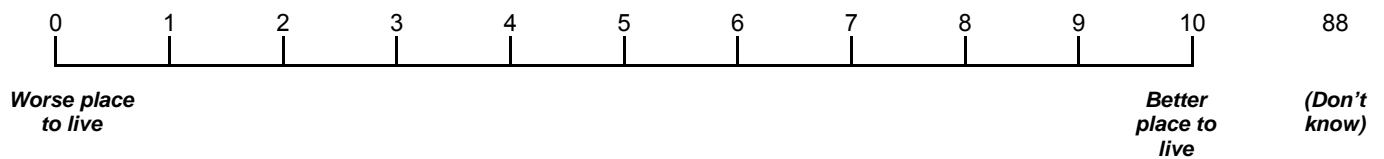
51. I am now going to read out some statements about men and women and their place in the family. Please tell me how much you agree or disagree with each one.

a. Firstly:

"A woman should be prepared to cut down on her paid work for the sake of her family"

Do you… **READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

74

b. Secondly:

"Men should take as much responsibility as women for the home and children"

Do you…**READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

c. Thirdly:

"When jobs are scarce, men should have more right to a job than women"

Do you…**READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

d. Finally:

"When there are children in the home, parents should stay together even if they don't get along"

Do you…**READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

52. I'm now going to read out two more statements. Please say to what extent you agree or disagree with each one.

a. Firstly:

"Gay men and lesbians should be free to live their own life as they wish"

Do you…**READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

b. Secondly:

"Whatever the circumstances, the law should always be obeyed"

Do you…**READ OUT**

…agree strongly,	1
agree,	2
neither agree nor disagree,	3
disagree,	4
or, disagree strongly	5
(Don't know)	8

53. Regardless of whether you belong to a particular religion, how religious would you say you are? Please use a scale from 0 to 10, where 0 means you are not at all religious and 10 means you are very religious. Pick a number from 0 to 10.

0	1	2	3	4	5	6	7	8	9	10	88
Not at all religious										*Very religious*	*(Don't know)*
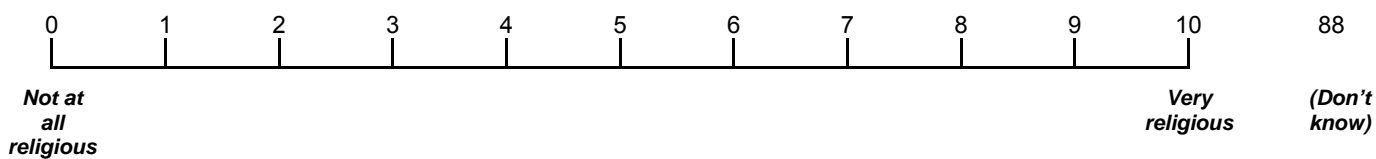
54. Apart from special occasions such as weddings and funerals, about how often do you attend religious services nowadays? Is it… **READ OUT**

…at least once a week,	1
at least once a month,	2
less often than that,	3
or never	4
(Don't know)	88

55. How would you describe the area where you live? Is it… **READ OUT**

…a big city,	1
the suburbs or outskirts of a big city,	2
a town or small city,	3
a country village,	4
or, a farm or home in the countryside	5
(Don't know)	8

[Country-specific question and codes for coding into ESS Coding Frame]

56. What is the highest level of education you have achieved? Is it… **READ OUT**

	00
	01
COUNTRY SPECIFIC QUALIFICATIONS (SEE F6/ CARD 46)	02
	03
	04
	05
	06
(Don't know)	88

76

57. Can I just check, did you do any paid work (of an hour or more) in the last seven days?

Yes	1	**ASK 25**
No	2	**GO TO 24**

58. Have you ever had a paid job?

Yes	1	**ASK 25**
No	2	**GO TO 28**

ASK IF 1 AT 23 OR 1 AT 24

INTERVIEWER: If respondent currently in work (1 at 23), ask 25 about <u>current</u> job; if not in paid work but had a job in the past (1 at 24), ask 25 about <u>last</u> job.

59. What is/ was the name or title of your main job? **WRITE IN**

_____

60. In your main job, what kind of work do/did you do most of the time? **WRITE IN**

_____

_____

_____

61. a.　　　So, just to be sure, is/was it… **READ OUT**

…non-manual,	1	**ASK 27b**
or manual?	2	**GO TO 27c**

IF 1 AT 27a

    b.   And how would you describe your job? Is/was it… **READ OUT**

INTERVIEWER: USE EXAMPLES GIVEN IF RESPONDENT ASKS FOR CLARIFICATION

…senior professional,	1

(e.g. accountant, solicitor, medical practitioner, scientist, civil/mechanical engineer)

other professional,	2

(e.g. teacher, nurse, physiotherapist, social worker, welfare officer, artist, musician, police officer [sergeant or above], software designer)

senior manager or administrator,	3

(usually responsible for planning, organising and co-ordinating work and for finance, such as finance manager, chief executive)

Middle or junior manager or administrator,	4

(e.g. office manager, retail manager, bank manager, restaurant manager, warehouse manager, publican)

or, other clerical?	5

(e.g. secretary, personal assistant, clerical worker, office clerk, call centre agent, nursing auxiliary, nursery nurse)

NOW ASK Q28

IF 2 AT 27a

    c.   And how would you describe your job? Is/was it…**READ OUT**

INTERVIEWER: USE EXAMPLES GIVEN IF RESPONDENT ASKS FOR CLARIFICATION

…a technical or craft occupation,	1

(e.g. motor mechanic, fitter, inspector, plumber, printer, tool maker, electrician, gardener, train driver)

An intermediate manual or service occupation,	2

(e.g. postal worker, machine operative, security guard, caretaker, farm worker, catering assistant, receptionist, sales assistant)

Or a routine manual or service occupation?,	3

(HGV driver, van driver, cleaner, porter, packer, sewing machinist, messenger, labourer, waiter/waitress, bar staff)

62. People's income comes from lots of different sources such as wages or salaries, pensions, social benefits, savings, investments and so on. Please give us an estimate of your household's total **net** income from **all** sources.

     WRITE IN _____

63. How long a period does that cover?

Week	1
Fortnight	2
Four weeks	3
Calendar month	4
Year	5
(Refused)	7
(Don't know)	8

64. **INTERVIEWER CODE SEX OF RESPONDENT**

Male	1
Female	2

65. In which year were you born?

     WRITE IN YEAR: | 1 | 9 | | |

66. Do you have access to the Internet at home or at work?

Yes, at home	1
Yes, at work	2
Yes, both at home and work	3
No	4

67. If you were asked to do a survey at home that would take about an hour, how would you choose to answer the questions? Would it be… **READ OUT**

…face-to-face interview,	1
telephone interview,	2
filling in a paper questionnaire,	3
filling in a questionnaire on the web,	4
or, some other way? **(WRITE IN)** _____	5

QUESTIONS FOR INTERVIEWERS (ALL GROUPS)

INTERVIEWER: PLEASE COMPLETE THE FOLLOWING QUESTIONS AT THE END OF THE INTERVIEW.

1. Did the respondent ask for clarification on any questions?

Never	1
Almost never	2
Now and then	3
Often	4
Very often	5
(Don't know)	8

2. Did you feel the respondent was reluctant to answer any questions?

Never	1
Almost never	2
Now and then	3
Often	4
Very often	5
(Don't know)	8

3. Did you feel that the respondent tried to answer the questions to the best of his or her ability?

Never	1
Almost never	2
Now and then	3
Often	4
Very often	5
(Don't know)	8

4. Overall, did you feel the respondent understood the questions?

Never	1
Almost never	2
Now and then	3
Often	4
Very often	5
(Don't know)	8

5. Did anyone else present influence the responses given?

Yes	1
No	2
(Don't know)	8

6. Did the respondent appear to be distracted in any way during the interview?

Yes	1
No	2
Maybe	3
(Don't know)	8

9.2 Questionnaire design

9.2.1 Question characteristics

Table A1: Question characteristics

Item	Topic	Factual or attitude	Show card number	Format (in aural modes if different)	Socially desirable responses (end point label)
Q1	Time watching TV	f	1	8 cats (open)	–
Q2	Time watching TV news	f	1	8 cats (open)	–
Q3	Trust people	a	2	11 pt scale	7-10 (most can be trusted)
Q4	Life satisfaction	a	3	11 pt scale	7-10 (extremely satisfied)
Q5	Political interest	a	–	4 pt scale	1-2 (very interested)
Q6	Political understanding	a	4	5 pt scale	1-2 (never too difficult)
Q7	Political opinion	a	5	5 pt scale	4-5 (very easy to form opinion)
Q8a	Trust institutions: parliament	a	6	11 pt scale	–
Q8b	Trust institutions: legal system	a	6	11 pt scale	–
Q8c	Trust institutions: police	a	6	11 pt scale	–
Q8d	Trust institutions: politicians	a	6	11 pt scale	–
Q8e	Trust institutions: parties	a	6	11 pt scale	–
Q8f	Trust institutions: EU parliament	a	6	11 pt scale	–
Q8g	Trust institutions: UN	a	6	11 pt scale	–
Q9	Voted last national election	f	–	yes/ no	1 (yes)
Q10	Party voted for	f	–	8 cats	–
Q11	Immigration: same ethnicity	a	7	4 pt scale	1-2 (allow many to come)
Q12	Immigration: different ethnicity	a	7	4 pt scale	1-2 (allow many to come)
Q13	Immigration: poor outside EU	a	7	4 pt scale	1-2 (allow many to come)
Q14	Immigration: impact on economy	a	8	11 pt scale	7-10 (good for economy)
Q15	Immigration: impact on culture	a	9	11 pt scale	7-10 (cultural life enriched)
Q16	Immigration: impact on living standards	a	10	11 pt scale	7-10 (better place to live)
Q17a	Gender role: mothers should not work	a	11	agree: 5 pt scale	1-2 (strongly agree)
Q17b	Gender role: men responsible for family	a	11	agree: 5 pt scale	1-2 (strongly agree)
Q17c	Gender role: men more right to jobs	a	11	agree: 5 pt scale	4-5 (strongly disagree)
Q17d	Gender role: parents should not divorce	a	11	agree: 5 pt scale	1-2 (strongly agree)
Q18a	Homosexuals free to live own lifestyle	a	11	agree: 5 pt scale	1-2 (strongly agree)
Q18b	Law should always be obeyed	a	11	agree: 5 pt scale	1-2 (strongly agree)
Q19	Religiosity	a	12	11 pt scale	6-10 (very religious)
Q20	Church attendance	f	13	7 cats (4)	1 (every day)
Q21	Area of residence	f	14	5 cats	–
Q22	Qualifications	f	15	15 cats	–
Q28	Household income	f	17	12 cats (open)	6-10 (>€ 1500/month)
Q31	Internet access	f	–	4 cats	–

Notes: The numeric items asked as open-ended questions in the aural modes (q1, q2, q28) were coded to correspond to the face-to-face showcard categories. For q20 the 7 showcard categories were collapsed to correspond to 4 the aural categories.

9.2.2 Relationship to ESS round 2 questionnaire

Mixed modes	ESS Round 2
1	A1
2	A2
3	A8
4	B24
5	B1
6	B2
7	B3
8a	B4
8b	B5
8c	B6
8d	B7
8e	B8
8f	B9
8g	B10
9	B11
10	B12
11	B35
12	B36
13	B37
14	B38
15	B39
16	B40
17a	G6
17b	G7
17c	G8
17d	G9
18a	B31

Mixed modes	ESS Round 2
18b	Not in R2
19	C13
20	C14
21	F5
22	F6
23	F9
24	F10
25	F22
26	F23
27	F54
28	F32
29	F2
30	F3
32	Not in ESS
33	Not in ESS
34	Not in ESS
35	Not in ESS
36	Not in ESS

Interviewer Questions	ESS Interviewer Questionnaire
1	J1
2	J2
3	J3
4	J4
5	Not in ESS
6	Not in ESS

9.3 Appendix tables

Table A2: Summary statistics

Variable	F2f showcard Obs	F2f showcard Mean	F2f showcard Std Dev	F2f no showcard Obs	F2f no showcard Mean	F2f no showcard Std Dev	Telephone Obs	Telephone Weight Obs	Telephone Mean	Telephone Std Dev	Overall Min	Overall Max
q1	515	4.56	2.03	512	4.70	1.99	685	1027	4.15	2.04	0	7
q2	499	2.08	1.44	500	2.31	1.25	666	999	2.34	1.32	0	7
q3	509	4.42	2.61	515	4.41	2.54	678	1018	4.69	2.46	0	10
q4	514	5.66	2.58	517	5.75	2.39	682	1024	5.73	2.44	0	10
occup4	493	4.45	2.63	511	4.60	2.81	670	1003	4.38	2.93	1	13
q5	512	2.56	0.98	516	2.54	1.01	680	1021	2.31	0.89	1	4
q6	494	2.82	1.28	491	2.85	1.25	652	986	2.62	1.28	1	5
q7	496	3.58	0.99	500	3.55	0.97	656	989	3.64	0.93	1	5
q8a	497	3.86	2.55	498	3.80	2.50	660	997	3.90	2.45	0	10
q8b	501	4.66	2.66	497	4.61	2.57	657	991	4.86	2.55	0	10
q8c	500	5.41	2.49	509	5.53	2.40	664	1003	5.46	2.39	0	10
q8d	492	2.93	2.21	502	2.79	2.23	659	993	3.03	2.17	0	10
q8e	486	2.94	2.22	496	2.78	2.20	640	966	3.01	2.24	0	10
q8f	422	5.17	2.57	409	4.99	2.40	531	817	5.08	2.55	0	10
q8g	431	5.38	2.67	414	5.16	2.51	533	833	5.39	2.57	0	10
voted	502	0.86	0.35	506	0.87	0.34	671	1007	0.89	0.32	0	1
q10	320	3.83	1.75	333	4.03	1.73	426	654	4.03	1.91	1	8
q11	492	2.46	0.98	485	2.45	1.02	642	968	2.17	0.92	1	4
q12	495	2.84	0.90	484	2.78	0.91	636	962	2.59	0.84	1	4
q13	489	3.02	0.89	485	2.99	0.88	628	948	2.88	0.86	1	4
q14	476	4.36	2.67	469	4.29	2.59	628	954	4.66	2.52	0	10
q15	479	5.54	2.87	471	5.82	2.70	644	974	6.28	2.65	0	10
q16	465	4.17	2.53	461	4.28	2.39	610	928	4.88	2.31	0	10
q17a	505	2.57	1.27	514	2.70	1.22	671	1010	2.86	1.22	1	5
q17b	511	1.37	0.74	518	1.40	0.73	682	1023	1.24	0.55	1	5
q17c	504	2.70	1.40	509	2.86	1.35	677	1016	2.96	1.30	1	5
q17d	493	3.56	1.31	505	3.45	1.24	669	1003	3.55	1.10	1	5
q18a	491	2.80	1.37	494	2.87	1.35	667	1000	2.76	1.29	1	5
q18b	508	1.58	0.85	516	1.59	0.81	684	1026	1.42	0.79	1	5
q19	511	4.25	3.45	512	3.94	3.39	678	1016	4.55	3.22	0	10
q20	504	3.13	0.97	503	3.25	0.89	680	1018	3.12	1.00	1	4
q21	514	1.13	0.36	518	1.16	0.41	685	1027	1.13	0.39	1	4
q22	511	4.02	1.56	514	3.96	1.47	684	1025	4.25	1.51	1	6
q28	396	3.69	1.30	331	5.40	1.51	524	783	5.65	1.83	1	12
q31	511	2.97	1.28	516	2.96	1.28	684	1026	2.98	1.20	1	4

Notes: Summary statistics are weighted but not adjusted for differences in sample composition.

Table A3: 11 Mode effects in 11 point semantic differential Scales?

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q3	0.1451	0.3485	0.0348	0.8280	0.1446	0.3329
q4	-0.0235	0.8674	0.1637	0.2753	-0.1375	0.3123
q8a	-0.0652	0.6812	-0.0011	0.9947	0.0161	0.9169
q8b	0.1401	0.3928	-0.0456	0.7882	0.1679	0.2927
q8c	0.0387	0.8007	0.1575	0.3174	-0.0520	0.7272
manual*mode	0.8076*	0.0333	0.4709	0.2174	0.4334	0.2318
q8d	0.0076	0.9554	-0.1295	0.3694	0.1702	0.2073
q8e	-0.0222	0.8760	-0.1476	0.3027	0.2050	0.1434
q8f	-0.0441	0.8009	-0.1376	0.4391	0.0738	0.6618
q8g	-0.0114	0.9488	-0.2319	0.2049	0.1926	0.2615
q14	0.2464	0.1356	-0.0822	0.6358	0.3415*	0.0343
male*mode	0.8106*	0.0204	0.4731	0.1854	0.3223	0.3407
q15	0.6631***	0.0001	0.2443	0.1778	0.4242*	0.0112
q16	0.5989***	0.0001	0.1285	0.4172	0.5120***	0.0006
q19	0.3041	0.1325	-0.3162	0.1350	0.6708***	0.0006

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A4: Mode effects in 5 point scales?

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q6	-0.1128	0.1416	0.0522	0.5011	-0.1528*	0.0413
manual*mode	0.5595**	0.0042	0.1246	0.5241	0.4339*	0.0195
q7	0.0300	0.6101	-0.0349	0.5664	0.0694	0.2177
age*mode	0.0523**	0.0095	0.0514*	0.0188	0.0000	0.9997
age2*mode	-0.0005*	0.0121	-0.0005*	0.0186	0.0000	0.9376
q17a	0.2777***	0.0004	0.1063	0.1807	0.1200	0.1082
q17b	-0.1329**	0.0014	0.0200	0.6743	-0.1691***	0.0000
q17c	0.1964*	0.0162	0.1917*	0.0246	-0.0044	0.9557
q17d	-0.0550	0.4516	-0.1127	0.1590	0.0800	0.2618
q18a	-0.0216	0.7941	0.0344	0.6876	-0.0591	0.4576
age2*mode	0.0005*	0.0487	0.0001	0.8272	0.0004	0.1468
q18b	-0.1875***	0.0002	0.0104	0.8455	-0.1951***	0.0000

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A5: Mode effects in 4 point scales?

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q5	-0.2104***	0.0002	-0.0107	0.8596	-0.2054***	0.0003
q11	-0.2894***	0.0000	-0.0173	0.7843	-0.2674***	0.0000
age*mode	-0.0477*	0.0286	-0.0372	0.1324	-0.0163	0.4876
age2*mode	0.0004*	0.0322	0.0003	0.1593	0.0002	0.4779
q12	-0.2429***	0.0000	-0.0656	0.2582	-0.1822***	0.0008
male*mode	-0.3194**	0.0064	-0.1549	0.2083	-0.1577	0.1787
q13	-0.1313*	0.0174	-0.0404	0.4828	-0.0908	0.0978

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A6: Mode effects in questions modified for aural administration?

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q1	-0.3629**	0.0024	0.0520	0.6761	-0.4508***	0.0001
q2	0.2896***	0.0006	0.2152*	0.0124	0.0834	0.2814
manual*mode	0.5549**	0.0046	0.2654	0.1843	0.3187	0.0840
q20	0.0170	0.7778	0.1236*	0.0381	-0.1249*	0.0292
q28_mid	729.4283***	0.0000	447.7694***	0.0000	297.9172***	0.0001
age*mode	42.5823*	0.0446	-13.9932	0.5147	57.3477*	0.0261
age2*mode	-0.4755**	0.0081	0.1409	0.4407	-0.6187**	0.0045
manual*mode	-251.5073*	0.0202	-56.4909	0.5644	-184.9476	0.1878

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A7: Mode effects in questions with other response categories?

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
q10	0.1359	0.3463	0.2343	0.0885	-0.0894	0.5239
age2*mode	0.0008	0.1116	0.0010*	0.0405	-0.0001	0.8965
q31	0.0803	0.2469	-0.0448	0.5269	0.1250	0.0640
q31 yes/no	0.2161	0.1876	0.1045	0.5255	0.0821	0.6209

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A8: Mode effects in yes/no questions

	F2f showcard vs. telephone		F2f showcard vs. f2f no showcard		F2f no showcard vs. telephone	
	Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
voted	0.0471	0.8079	0.0485	0.8063	0.0044	0.9817

Notes: * p<0.05, ** p<0.01, *** p<0.001.

Table A9: Time spent watching television (q1)

	fixed	f2f_sc	f2f_nosc	Total
0	0.0273	0.0272	0.0234	0.0263
<½ hr	0.0224	0.0524	0.0137	0.0278
½-1 hr	0.2580	0.1282	0.1680	0.2030
1-1½ hrs	0.0896	0.0738	0.0664	0.0798
1½-2 hrs	0.2103	0.2233	0.2168	0.2152
2-2½ hrs	0.0662	0.0738	0.0391	0.0613
2½-3hrs	0.1149	0.1864	0.2168	0.1582
>3 hrs	0.2113	0.2350	0.2559	0.2283

Note: not adjusted for differences in sample composition.

Table A10: Time spent watching news programmes on television (q2)

all modes	fixed	f2f_sc	f2f_nosc	Total
0	0.0531	0.0461	0.0460	0.0495
<½ hr	0.1101	0.3407	0.0900	0.1627
½-1 hr	0.5746	0.3707	0.6360	0.5390
1-1½ hrs	0.0911	0.1082	0.0660	0.0891
1½-2 hrs	0.1091	0.0701	0.1160	0.1011
2-2½ hrs	0.0170	0.0140	0.0020	0.0125
2½-3hrs	0.0290	0.0240	0.0280	0.0275
>3 hrs	0.0160	0.0261	0.0160	0.0185

Note: not adjusted for differences in sample composition.

Table A11: Relationship between variables: response distribution

		Telephone vs. f2f showcard		F2f no showcard vs. f2f showcard		Telephone vs. f2f no showcard	
		Coeff.	P-Value	Coeff.	P-Value	Coeff.	P-Value
Q1_all	Mode	0.2749	0.4774	0.6377	0.2002	-0.4640	0.3536
	Income	0.0524	0.5247	0.0040	0.9596	-0.1055	0.2139
	Income*mode	-0.1237	0.1531	-0.1079	0.2783	0.0123	0.8884
Q2_all	Mode	-0.2111	0.6393	-0.1240	0.7826	-0.0784	0.8553
	Political interest score	-0.2377***	0.0000	-0.2095***	0.0000	-0.1682***	0.0001
	Pol. Interest*mode	**0.1184***	0.0338	0.0943	0.0908	0.0253	0.6424
Q5	Mode	-0.6220	0.0810	-0.4238	0.1899	-0.1422	0.6698
	Voted	-1.2331***	0.0000	-1.1315***	0.0000	-0.7984**	0.0022
	Voted*mode	0.1937	0.6055	0.4060	0.2415	-0.2652	0.4537
Q5	Mode	-0.1500	0.8213	-0.0603	0.9242	0.1250	0.8565
	Gender role score	-0.0235	0.4896	-0.0149	0.6185	-0.0132	0.7352
	Gender role*mode	-0.0198	0.6683	0.0038	0.9334	-0.0400	0.4106
Q11	Mode	-0.4939**	0.0039	-0.0514	0.7602	-0.4425**	0.0096
	In work	-0.0852	0.6819	-0.0392	0.8345	0.0214	0.9205
	In work*mode	-0.1130	0.6379	0.0740	0.7589	-0.1512	0.5341
Q12	Mode	-0.4828*	0.0111	-0.1445	0.4213	-0.3211	0.0684
	In work	0.1889	0.4075	0.1681	0.4075	0.3379	0.1201
	In work*mode	-0.1721	0.5043	0.0283	0.9086	-0.2254	0.3725
Q13	Mode	-0.3778*	0.0367	-0.2650	0.1383	-0.0933	0.5782
	In work	-0.1568	0.4541	-0.1881	0.3279	0.2572	0.2048

	In work*mode	0.1175	0.6319	0.3317	0.1758	-0.2420	0.3179
Q15	Mode	0.2942	0.0647	0.1601	0.3047	0.1415	0.3728
	In work	-0.2343	0.2480	-0.1600	0.4127	-0.3230	0.1076
	In work*mode	0.2430	0.2943	-0.0400	0.8665	0.3001	0.1910
Q16	Mode	0.4385*	0.0120	0.2432	0.1480	0.2465	0.1496
	In work	-0.1416	0.4880	-0.0691	0.7207	-0.4515*	0.0303
	In work*mode	0.1204	0.6090	-0.2582	0.2874	0.3681	0.1172
Q17a	Mode	0.4186**	0.0085	0.1439	0.3838	0.1879	0.2249
	In work	0.2729	0.1796	0.1707	0.4081	0.2849	0.1129
	In work*mode	0.0223	0.9219	0.0372	0.8727	0.0403	0.8538
Q17a	Mode	0.2991	0.1245	-0.0184	0.9199	0.2271	0.2185
	Religiosity	-0.0799**	0.0027	-0.0794**	0.0030	-0.0520	0.0605
	Religiosity*mode	0.0295	0.4068	0.0352	0.3411	0.0022	0.9529
Q17b	Mode	-0.2294	0.2789	0.3451	0.0933	-0.6310**	0.0020
	In work	-0.0724	0.7709	0.0567	0.8120	-0.5369*	0.0227
	In work*mode	-0.3587	0.2364	-0.4165	0.1481	0.0937	0.7536
Q17b	Mode	-0.3663	0.1363	0.2383	0.2919	-0.6772**	0.0042
	Religiosity	-0.0073	0.8265	-0.0157	0.6215	-0.0512	0.1167
	Religiosity*mode	-0.0074	0.8692	-0.0334	0.4452	0.0328	0.4578
Q17c	Mode	0.2348	0.1459	0.2872	0.0738	-0.0736	0.6369
	In work	0.1277	0.5325	0.1629	0.4090	0.1685	0.3761
	In work*mode	0.0885	0.7013	-0.0150	0.9492	0.1235	0.5860
Q17c	Mode	0.3257	0.1141	0.2772	0.1371	-0.0130	0.9439
	Religiosity	-0.0598*	0.0314	-0.0654*	0.0148	-0.0832**	0.0018
	Religiosity*mode	-0.0107	0.7690	-0.0121	0.7354	0.0111	0.7564
Q18b	Mode	-0.4698	0.2748	0.1158	0.7704	-0.5269	0.2215
	Trust institutions sc.	-0.0104	0.2241	-0.0109	0.2020	-0.0135	0.1002
	Trust*mode	-0.0062	0.6226	-0.0037	0.7538	-0.0051	0.6960
Q18b	Mode	-0.3898	0.0801	0.3757	0.0660	-0.7436***	0.0005
	Religiosity	0.0239	0.3847	0.0236	0.4016	-0.0400	0.1356
	Religiosity*mode	-0.0506	0.2026	-0.0762	0.0502	0.0199	0.6126
Q28_all	Mode	2.9519***	0.0000	3.5836***	0.0000	-0.4510	0.2389
	Life satisfaction	0.1166**	0.0050	0.1385**	0.0012	0.0937	0.0584
	Satisfaction*mode	0.0235	0.6615	-0.0524	0.4087	0.0824	0.1853

Notes: * p<0.05, ** p<0.01, *** p<0.001. Proportional odds models including socio-demographics in addition to the covariates reported.