

Causes of Mode Effects: Separating out Interviewer and Stimulus Effects in Comparisons of Face-to-Face and Telephone Surveys

Caroline Roberts¹, Annette Jäckle and Peter Lynn²
City University¹
University of Essex²

Abstract

We identify the causes of mode effects in comparisons of face-to-face and telephone surveys, by testing for differences in the extent of satisficing and social desirability bias due to differences in the stimulus (visual vs. aural presentation of response options) and the presence vs. absence of the interviewer. The stimulus did not lead to differential measurement error; the presence or absence of the interviewer however did. Controlling for differences in the stimulus between both modes, telephone respondents were far more likely to give socially desirable responses than face-to-face respondents.

Keywords: measurement error, CATI, showcards, visual, aural, satisficing, social desirability

1. Introduction

Although previous studies have tested differences in measurement 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).

The principal differences between face-to-face and telephone interviewing are the *sensory channel* available for the transmission of information and the *physical presence* of the interviewer (de Leeuw 2005). Face-to-face interaction, for example, 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

rapprochement between interviewers and respondents in face-to-face interviews, making in-person interaction more intimate than that conducted by telephone. Face-to-face respondents are also 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. Holbrook et al. (2003) found evidence that such differences between the two modes lead to variation in the extent of respondent satisficing (the shortcutting of the response process, either due to differences in the task difficulty or lack of respondent motivation, see Krosnick 1991) and social desirability bias. Their study however ignored issues of stimulus comparability by only examining questions for which no showcards were used.

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. 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. 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 (e.g. Sykes and Collins 1988) and serve to distract them.

Our experiment 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 attributable to 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.

The study is part of an ongoing methodological programme of research designed to inform decisions

by the European Social Survey (ESS) on whether to change the current policy of single-mode data collection using face-to-face interviews to a mixed-mode data collection strategy, and if so, which modes to mix and how (for a discussion of the issues faced by the ESS, see Jäckle, Roberts and Lynn 2006.).

2. Hypotheses

Modes are likely to lead to differences in measurement, 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. 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.

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 (DeMaio 1984).

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.

To assess the likely causes of mode effects, we tested a number of potential explanations of how the differences between telephone and face-to-face interviewing 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 Tourangeau,

Rips, and Rasinski (2000), Holbrook, Green and Krosnick (2003) and de Leeuw (2005).

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.

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. 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.

H3: The interviewer's presence reduces 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 or 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).

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. In comparison, telephone respondents may feel less confident that the interviewer will protect the confidentiality of their responses and may consequently be less willing to disclose personal information. If this is true then we should see more socially desirable answers with telephone interviewing.

We can only test for the combined effect of the interviewer's presence on task difficulty (H3) and respondent motivation (H4) and similarly on anonymity (H6) and rapport (H7). The distinction between these potential mechanisms is nonetheless important, since methods of minimising mode effects would have to be quite different depending on the primary cause. Should we find evidence consistent with these hypotheses, further research would therefore be necessary to disentangle the causes.

To test these potential explanations for mode differences, we compare the extent of respondent satisficing and socially desirable 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 socially desirable 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 null hypotheses that differences in satisficing and social desirability bias are not due to mode cannot be rejected.

3. Method

3.1 Research Design

Our objective was to compare data from face-to-face survey interviews with data from telephone interviews, controlling for how questions are asked in each mode. 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) .

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:

- Group 1 Face-to-face interview with showcards
- Group 2 Face-to-face interview without showcards
- Group 3 Telephone interview

3.2. Sampling and Response Rates

Fieldwork was conducted by Gallup Europe in the Greater Budapest region of Hungary starting in July 2005. Focusing on the capital area 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. In total, 515 respondents were interviewed face-to-face using showcards, 518 respondents were interviewed face-to-face without showcards and 887 were interviewed over the telephone. The response rate for the telephone group was 32% and that for the face-to-face groups combined was 33% (AAPOR response rate 1).

Compared to the face-to-face samples, 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 differences in sample composition, by including controls for socio-demographics in the multivariate models. The motivation for this approach was to avoid confounding differential non-response with differential measurement error, though we recognise that some differential non-response may remain even after the socio-demographic controls.

3.3. Questionnaire

The interviews consisted of a subset of questions from the core questionnaire of the European Social Survey. For group 1, the question and showcard design was essentially identical to the ESS 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 were selected to provide a variety of indicators of data quality across each of the two modes, based on those used in other mode comparison studies. A further criterion was to select items that were believed to be most likely to be susceptible to mode effects, so that we could 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 table 2 for question wording). In addition, the questionnaire included socio-demographic measures (including sex, age, education, occupation and income).

3.4 Outcome Measures

We used the following indicators of respondent satisficing and social desirability bias:¹

1. *Non-differentiation* occurs when respondents choose a response category that seems appropriate for the first item of a battery of questions measured on the same scale and stick to that response for all other items in the battery (Locander, Sudman and Bradburn 1976). We measured the extent of non-differentiation using answers to two batteries of scale questions. 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 both scales.

2. *Acquiescence* refers to the tendency to agree with or accept any assertion, regardless of its content (Couch and Keniston 1960). We used answers to agree/disagree questions to calculate the proportion of 'agree' responses (excluding 'strongly agree') 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.

3. *Response order effects* result from 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. 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, including only those 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 first-category responses across the 12 items. At the item level we used binary indicators which take the

¹ Krosnik (1991) identifies additional indicators of satisficing, such as endorsing the status quo instead of social change or randomly choosing response alternatives.

value 1 if the first (last) category was chosen and 0 otherwise.

4. *Social desirability bias* refers to responses being more socially desirable than the true answer would be. This may be the result of deliberate editing of responses, or impression management, to make the respondent appear in a better light. We did not test the social desirability connotations of items, as recommended by Holbrook et al. (2003), but selected 23 items, some of which have been shown elsewhere to have social desirability connotations (the selected categories are specified in the final column of Table 2). We created an overall indicator using the proportion of answers to the 23 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. *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 deemed to be socially undesirable and 0 otherwise.

4. Results

We first examined the extent of mode effects in the data, before evaluating the hypotheses about the potential causes of mode effects. All models described here included controls for differences in sample composition (age, age squared, male, educational qualification and occupation).

We examined the magnitude of mode effects by comparing the similarity of response distributions across 33 items in the questionnaire. Two methods were used: 1) an OLS 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. For a discussion of these methods and the tables of results, see Jäckle, Roberts and Lynn (2006). Of the 33 items tested, just 13 were affected at the mean 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, but revealed mode effects that were not visible at the mean for a further 4 items.

To evaluate the hypotheses of the causes of mode effects, we estimated models for each of the indicators of satisficing and social desirability bias. Separate models were run for the comparison of the two face-to-face modes and for the comparison of the two aural modes. Based on these models, Table 1 reports the differences in predicted probabilities of satisficing and social disability bias. For each of the five indicators, we compared the predicted probability of that type of response for a sample person of mean age (57) and modal characteristics (female, finished high school or equivalent and in a clerical occupation), comparing the two face-to-face modes (to identify stimulus effects) and the two aural modes (to identify interviewer effects).

3.2.1. Satisficing

There was no evidence that showcards decreased the extent of satisficing (H1). The probability of non-differentiation (not reported) was no different between the two face-to-face groups and neither was the probability of acquiescence.

There was weak evidence that showcards led to more primacy responses (around 6% more likely for Q17a and Q17c). If the differences were due to the showcards then there should not have been any differences between the two aural groups. Instead, telephone respondents were overall 2.9% more likely to give primacy responses than the face-to-face no showcard group and 3.3% less likely to give recency responses. This suggests that the observed differences in the reporting of extreme responses were not due to the showcards.

There was no evidence that telephone respondents were more likely to satisfice (H3/H4), judging by either the non-differentiation or acquiescence measures. In fact overall the probability of satisficing was 2.8% lower with telephone than face-to-face without showcards. There was also no evidence that lower ability respondents were more likely to satisfice than higher ability respondents (H5). The interactions of mode and ability were not significant at conventional levels in any of the models.

3.2.2. Social Desirability

Although we did not find differences between modes in the extent of satisficing, we did find strong differences in social desirability bias (H6/H7). Telephone respondents were more likely to give socially desirable answers (+3.5%) and less likely to choose undesirable extremes (-2.1%). In each case, there were significant and consistent differences at the item level for more than a third of items tested, ranging from 2-13%. This suggests that the differences between groups in the propensity to select first- and last-category responses, which could not be attributed to the use of showcards, can be attributed to social desirability bias leading to less reporting of socially undesirable extremes.

5. Discussion

The main differences in responses between modes in this study were due to social desirability bias associated with the physical distance between the interviewer and the respondent. There was no evidence that the interviewer's presence or the use of showcards affected the extent of satisficing. Telephone respondents in our study might however have been more willing to expend the necessary effort for two reasons: response rates were comparatively low, and therefore it is possible that sample members who did take part were more co-operative than the non-respondents would have been. In addition, the questionnaire was not only relatively short but also quite varied and therefore possibly more stimulating compared to longer surveys handling fewer topics.

The results of our research provide compelling further evidence supporting 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. Understanding more about how this type of bias operates and, in particular, about the cognitive mechanisms underlying it represents an important area for further research.

Acknowledgements

The study reported here forms part of a wider programme of methodological research on mixed mode data collection co-funded by Gallup Europe and the European Commission as part of the central funding of the European Social Survey. The authors wish to thank Robert Manchin and Agnes Illyes, who directed the fieldwork at Gallup Europe; Roger Jowell, director of the European Social Survey; Norman

Bradburn, Robert Groves, Willem Saris and members of the ESS Methods Group for advice and comments on the design of the experiment; Patrick Sturgis and Nick Allum for advice on statistical analysis and comments on early drafts.

References

- Couch, A. and Keniston, K. (1960) 'Yeasayers and Naysayers: Agreeing Response Set as a Personality Variable', *Journal of Abnormal and Social Psychology*, 60: 151-174.
- de Leeuw, E. D. (2005) 'To Mix or Not to Mix Data Collection Modes in Surveys', *Journal of Official Statistics*, 21(2): 233-255.
- DeMaio, T. J. (1984) 'Social Desirability and Survey Measurement: A Review', in C. F. Turner and E. Martin (eds) *Surveying Subjective Phenomena*, New York: Sage.
- Holbrook, A. L., Green, M. C. and Krosnick, J. A. (2003) 'Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias', *Public Opinion Quarterly*, 67(1): 79-125.
- Jäckle, A., Roberts, C. and Lynn, P. (2006) 'Telephone versus Face-to-Face Interviewing: Mode Effects on Data Quality and Likely Causes. Report on Phase II of the ESS-Gallup Mixed Mode Methodology Project', *ISER Working Paper, No. 2006-41*, Colchester: University of Essex.
- Krosnick, J. (1991) 'Response Strategies for Coping With the Cognitive Demands of Attitude Measures in Surveys', *Applied Cognitive Psychology*, 5: 213-236.
- Krosnick, J. A. and Alwin, D. F. (1987) 'An Evaluation of a Cognitive Theory of Response Order Effects in Survey Measurement', *Public Opinion Quarterly*, 51: 201-219.
- Locander, W., Sudman, S. and Bradburn, N. (1976) 'An Investigation of Interview Method, Threat, and Response Distortion', *Journal of the American Statistical Association*, 71: 269-275.
- Rogers, T. F. (1976) 'Interviews by Telephone and in Person: Quality of Responses and Field Performance', *Public Opinion Quarterly*, 40: 51-56.
- Sykes, W. and Collins, M. (1988) 'Effects of mode of interview: Experiments in the UK', in R. M. Groves, et al. (eds) *Telephone Survey Methodology*, New York: Wiley.
- Tourangeau, R., Rips, L. J. and Rasinski, K. (2000) *The Psychology of Survey Response*, Cambridge: Cambridge University Press.

Table 1: Mode difference in predicted probability of satisficing and socially desirable reporting

| Item | Prob of Acquiescence | | Prob of Primacy | | Prob of Recency | | Prob of social desirability | | Prob undesirable extremes | |
|----------------------|-----------------------|--------------------|-----------------------|--------------------|-----------------------|--------------------|-----------------------------|--------------------|---------------------------|--------------------|
| | F2F no sc – F2F sc | Tel – F2F no sc | F2F no sc – F2F sc | Tel – F2F no sc | F2F no sc – F2F sc | Tel – F2F no sc | F2F no sc – F2F sc | Tel – F2F no sc | F2F no sc – F2F sc | Tel – F2F no sc |
| Overall ¹ | 0.001 | -0.028 * | -0.018 | 0.029 ** | -0.014 | -0.033 *** | 0.000 | 0.035*** | -0.012 | -0.021 *** |
| 1 | – | – | – | – | – | – | -0.006 | 0.117 *** | – | – |
| 2 | – | – | – | – | – | – | -0.037 | -0.013 | – | – |
| 3 | – | – | – | – | – | – | 0.016 | -0.034 | -0.011 | 0.021 |
| 4 | – | – | – | – | – | – | 0.031 | -0.043 | -0.008 | 0.007 |
| 5 | – | – | 0.026 | -0.002 | -0.002 | -0.101 *** | -0.019 | 0.129 *** | -0.002 | -0.101 *** |
| 6 | – | – | -0.015 | 0.048 * | -0.008 | 0.006 | -0.023 | 0.060 | -0.008 | 0.006 |
| 7 | – | – | -0.004 | -0.016 * | -0.021 | 0.007 | -0.030 | 0.053 | -0.004 | -0.016 * |
| 9 | – | – | – | – | – | – | 0.005 | 0.007 | -0.005 | 0.001 |
| 11 | – | – | 0.017 | 0.049 * | 0.007 | -0.127 *** | -0.014 | 0.109 *** | 0.007 | -0.127 *** |
| 12 | – | – | 0.006 | 0.016 | -0.030 | -0.127 *** | 0.014 | 0.045 | -0.030 | -0.127 *** |
| 13 | – | – | 0.001 | 0.005 | -0.035 | -0.063 * | -0.000 | 0.015 | -0.035 | -0.063 * |
| 14 | – | – | – | – | – | – | -0.016 | 0.028 | -0.011 | 0.033 |
| 15 | – | – | – | – | – | – | 0.013 | 0.067 * | -0.012 | 0.008 |
| 16 | – | – | – | – | – | – | 0.010 | 0.046 * | -0.022 | 0.004 |
| 17a | 0.022 | -0.032 | -0.058 * | -0.016 | 0.011 | -0.025 | -0.038 | -0.049 | 0.011 | -0.025 |
| 17b | 0.011 | -0.015 | -0.022 | 0.067 *** | -0.003 | -0.012 | -0.011 | 0.046 *** | -0.003 | -0.012 |
| 17c | -0.012 | -0.012 | -0.071 ** | 0.000 | 0.009 | -0.032 | 0.011 | 0.029 | -0.071 ** | 0.000 |
| 17d | -0.010 | 0.005 | -0.002 | -0.043 ** | -0.081 ** | -0.055 | -0.012 | -0.026 | -0.081 ** | -0.055 |
| 18a | 0.019 | 0.025 | -0.032 | 0.003 | -0.018 | -0.011 | -0.010 | 0.026 | -0.018 | -0.011 |
| 18b | 0.028 | -0.124 *** | -0.026 | 0.166 *** | -0.003 | – | 0.004 | 0.029 | -0.003 | – |
| 19 | – | – | – | – | – | – | -0.041 | 0.072 * | 0.012 | -0.066 * |
| 20 | – | – | – | – | – | – | -0.042 | 0.048 * | 0.042 | -0.031 |
| 28 | – | – | – | – | – | – | 0.144 *** | 0.069 ** | – | – |

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

¹ The ‘overall’ row shows the marginal effect of mode from OLS models of the proportion of items with satisficing or socially desirable responses. The remainder rows show differences between modes in the predicted probability of an answer corresponding to an indicator of satisficing or not, based on logit models. Predicted probabilities were calculated for respondents with mean age (57) and modal characteristics (women, finished high school or equivalent and other clerical occupation).

Table 2: Summary of Questions

| Item | Question Wording | Show card | Format (aural if different) | Socially desirable |
|------|--|-----------|-----------------------------|--------------------|
| 1 | On an average weekday, how much time, in total, do you spend watching television? | Y | 8 categories (open) | 1-3 |
| 2 | How much of this time is spent watching news or programmes about politics and current affairs? | Y | 8 categories (open) | 1-3 |
| 3 | Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? | Y | 11-pt. scale | 7-10 |
| 4 | All things considered, how satisfied are you with your life as a whole nowadays? | Y | 11-pt. scale | 7-10 |
| 5 | How interested would you say you are in politics? | N | 4-pt. scale | 1-2 |
| 6 | How often does politics seem so complicated that you can't really understand what is going on? | Y | 5-pt. scale | 1-2 |
| 7 | How easy or difficult do you find it to make your mind up about political issues? | Y | 5-pt. scale | 4-5 |
| 8a | On a scale of 0-10, how much do you personally trust each of these institutions: Hungary's parliament? | Y | 11-pt. scale | - |
| 8b | Trust institutions: the legal system? | Y | 11-pt. scale | - |
| 8c | Trust institutions: the police? | Y | 11-pt. scale | - |
| 8d | Trust institutions: politicians? | Y | 11-pt. scale | - |
| 8e | Trust institutions: political parties? | Y | 11-pt. scale | - |
| 8f | Trust institutions: the European parliament? | Y | 11-pt. scale | - |
| 8g | Trust institutions: the United Nations? | Y | 11-pt. scale | - |
| 9 | Did you vote in the last Hungarian national election? | N | yes/ no | yes |
| 11 | To what extent do you think Hungary should allow people of the same race or ethnic group as most Hungary's people to come and live here? | Y | 4-pt. scale | 1-2 |
| 12 | How about people of a different race or ethnic group? | Y | 4-pt. scale | 1-2 |
| 13 | How about people from poorer countries outside Europe? | Y | 4-pt. scale | 1-2 |
| 14 | Would you say it is generally bad or good for Hungary's economy that people come to live here from other countries? | Y | 11-pt scale | 7-10 |
| 15 | Would you say that Hungary's cultural life is generally undermined or enriched by people coming to live here from other countries? | Y | 11-pt scale | 7-10 |
| 16 | Is Hungary made a worse or a better place to live by people coming to live here form other countries? | Y | 11-pt scale | 7-10 |
| 17a | "A woman should be prepared to cut down on her paid work for the sake of her family." | Y | agree: 5-pt. scale | 1-2 |
| 17b | "Men should take as much responsibility as women for the home and children." | Y | agree: 5-pt. scale | 1-2 |
| 17c | "When jobs are scarce, men should have more right to a job than women." | Y | agree: 5-pt. scale | 4-5 |
| 17d | "When there are children in the home, parents should stay together even if they don't get along." | Y | agree: 5-pt. scale | 1-2 |
| 18a | "Gay men and lesbians should be free to live their own life as they wish." | Y | agree: 5-pt. scale | 1-2 |
| 18b | "Whatever the circumstances, the law should always be obeyed." | Y | agree: 5-pt. scale | 1-2 |
| 19 | Regardless of whether you belong to a particular religion, how religious would you say you are? | Y | 11-pt. scale | 6-10 |
| 20 | Apart from special occasions such as weddings and funerals, about how often do you attend religious services nowadays? | Y | 7 categories (4) | 1 |
| 28 | People's income comes from lots of different sources...Using this card, if you add up income from all sources, which letter describes your household's net household income? | Y | 12 categories (open) | 6-10 |