ISER Working Paper Series

Effects of Visual and Aural Communication of Categorical Response Options on Answers to Survey Questions

Peter Lynn
Institute for Social and Economic Research
University of Essex

Steven Hope
University College London

Annette Jäckle
Institute for Social and Economic Research
University of Essex

Pamela Campanelli
The Survey Coach

Gerry Nicolaas
NatCen Social Research

No. 2012-21
October 2012

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

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

www.iser.essex.ac.uk

Non-Technical Summary

The way that people answer questions can depend on whether they have had to listen to the question being read out or have had to read it for themselves. Both of these ways of asking questions are used in surveys. And sometimes surveys require the people taking part to *both* listen *and* read, for example when an interviewer reads out the question but asks the respondent to pick their answer from those listed on a card.

These different ways of being asked a question make a difference to how we think about the question. As a result, some people may be more likely to fully understand a question if it is asked in one way rather than another. Other people may be more likely to give a thoughtful, considered answer rather than a hasty unconsidered reaction if the question is asked in one way rather than another.

In this paper we investigate the effect on the answers received to survey questions of presenting response options visually rather than relying on the respondent hearing them read out by an interviewer. We try to identify the average effect, across a sample of people taking part in a survey, and also to identify whether particular types of people, such as those with greater mental capacity, are affected differently from others. We do this for several different types of survey questions, including questions about behaviour, attitudes and satisfaction.

For several survey questions we find evidence that presenting response options visually or orally does indeed make a difference to the answers obtained. Furthermore, we find that this effect is not uniform across all the respondents in the sample. Rather, it depends on the respondent's mental ability and on how motivated they are to answer the questions to the best of their ability. We also identify a need for further research as there are some aspects of the process of asking survey questions about which we still do not know very much.

Effects of visual and aural communication of categorical response options on answers to survey questions

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

Steven Hope
University College London

Annette Jäckle
Institute for Social and Economic Research, University of Essex

Pamela Campanelli
The Survey Coach

Gerry Nicolaas
NatCen Social Research

Abstract: Whether questions and answers are transmitted between interviewer and respondent by visual or aural communication can affect the responses given. We hypothesise that communication channel can affect either the respondent's understanding of the question or the tendency to satisfice. These effects may be mediated by the cognitive ability and motivation of the respondent and by the type and difficulty of the question. We test our hypotheses using data from a large-scale controlled experiment. We find support for the notion that visual presentation improves understanding of the question and reduces the tendency to satisfice. We also find that effects are stronger for respondents of higher cognitive ability and for motivated respondents.

Keywords: cognitive ability, end-labelled scales, primacy effect, recency effect, respondent motivation, response order, satisficing

JEL Codes: C81, C83

Acknowledgments: This research was funded by the Survey Design and Measurement initiative of the UK Economic and Social Research Council (ESRC) via a grant for the project "Mixed modes and measurement error", principal investigator Gerry Nicolaas (award no. RES-175-25-0007). We are grateful for input to the early stages of this project from our colleagues Rebecca Taylor, Alita Nandi and Nayantara Dutt. Preliminary versions of this paper were presented at the conference of the European Survey Research Association, Lausanne, July 2011, at a meeting of the social statistics section of the Royal Statistical Society, London, October 2011, and at a seminar at the University of Essex, February 2012. We acknowledge the useful comments of participants at these events.

Contact: plynn@essex.ac.uk

1. Introduction

Several decades ago, survey researchers became aware that the responses provided to survey questions could differ depending on the mode of administration of the survey. Early research documented the nature of the differences in responses between data collection modes (Payne 1951, Sudman & Bradburn 1974), but soon researchers wanted to go further and to understand the reasons for these differences. One possible explanation was that the way people process information depends on whether that information is received through visual or auditory channels (Duncan 1969). In this article we are concerned with the role in the question-answering process of two channels of communication, visual and oral/aural. We focus on survey items with categorical response options, either nominal or ordinal.

Previous research into the effects of communication channel on responses to such items has shown that aural communication of response options can tend to produce recency effects, whereby respondents have an increased tendency to choose (one of) the last option(s) presented, while visual communication can tend to produce primacy effects, an increased tendency to choose (one of) the first option(s) (Krosnick & Alwin, 1987), though observation of such effects is far from universal. These effects are generally believed to represent a reduction in the accuracy of responses, as the effects would be absent if respondents processed fully and equally all response options. The effects are therefore a manifestation of satisficing (Krosnick, 1991). Whether or not primacy and recency effects are observed may depend on factors such as the time and effort that would be required to fully process all options (Schwarz et al, 1992) the degree of uncertainty the respondent has regarding their true state, and the degree of ambiguity regarding the distinction between response categories. However, the role of these factors and others in influencing response-order effects is not well understood and a number of inconsistent findings from previous research studies have not been explained. Dillman et al (2009, p.316) conclude that "…*much remains to be learned about the conditions under which primacy and recency occur in surveys, some aspects of which may be entangled with other effects of visual versus aural communication of survey questions*."

Visual display of information may make the task of answering a survey question easier for survey respondents (Schwarz et al 1991). Visual display of response options provides cues and allows respondents to review the options at their own speed, in the order they choose, and

to re-read options. If respondents take full advantage of these opportunities, then they should be better able to understand the response task as a result.

We extend previous research in a number of ways. We acknowledge that the effect of different channels of communication – on understanding of the question and on the tendency to satisfice – could vary between survey respondents, depending on their cognitive ability and their motivation to make cognitive effort. This is consistent with the idea that ability to fully process information of any particular type depends on cognitive capacity (Dickens 2008) and with the idea that respondent satisficing depends on an interaction between motivation, ability and question difficulty (Krosnick, 1991). Therefore, we test for interactions of each effect of channel of communication with cognitive ability and motivation. Also, our study includes questions of different types (factual, attitudinal, satisfaction) and different levels of difficulty, to allow findings to be contrasted between questions with these different characteristics.

Furthermore, we note that there are three main communication components of the question-answering process: communicating the question to the respondent, communicating the response options or response format, and communicating the answer to the interviewer or survey researcher. The first two components involve the transmission of information from researcher/interviewer to respondent, while the third involves transmission in the opposite direction. Each component relies on one main channel of communication, but may also have a secondary channel. The channel need not be the same for the different components of the same question. For example, in a face-to-face interview using showcards, the respondent typically relies on aural perception to hear the question, visual perception to see the response options, and oral communication to report his or her answer.

A comparison of survey modes does not therefore correspond to a simple comparison of two channels of communication. Defining mode at the level of the survey hides important variation in channels of communication at the question and sub-question level. This is a major limitation of much previous research into survey mode effects, particularly where face-to-face interviewing is one of the modes studied. Several studies have treated channel of communication as a feature of the survey rather than of a specific question or question component. We therefore study the effects of channels of communication for specific components, or combinations of components, of the question-answering process. Notably, one part of our study varies the channel of communication of response options while holding

constant the channel of communication of the question stem, whereas another part varies both simultaneously, for the same questions.

Additionally, we study the effect of visual presentation of end-labelled response scales as addressed previously by Christian and Dillman (2004). But we overcome one limitation of that previous study, namely that the directionality of category numbering was counter-intuitive, potentially introducing extra respondent confusion.

We first set out our hypotheses regarding the role of channel of communication (section 2). We then introduce the data and methods that we use to address these hypotheses (section 3) before presenting our results (sections 4 to 7) and discussing the implications of our findings (section 8).

2. General Hypotheses

Aural perception is a basic skill important to almost all dimensions of daily life. Excepting a small minority of people with serious auditory impairments, there is believed to be only modest variation in ability to hear and comprehend sounds such as simple speech. However, comprehension of speech becomes more difficult as the number of words, complexity of words and complexity of language increases (Just et al. 1996; Jobard et al. 2007). Specifically, ability to process a question depends on whether the listener can hold in short-term memory all the relevant aspects of the question. Once a survey respondent has heard a particular word or phrase forming part of question, he or she will only be able to refer to that word or phrasing while formulating an answer if it has been retained in memory. *In extremis*, a respondent may ask the interviewer to repeat the question, or the response options, but this happens relatively infrequently[1]. It is likely that more commonly the respondent will impute the missing information, will ignore it on the assumption that it was unimportant, or will simply be unaware that some information is missing (Kellogg 2007, pp. 75). In consequence we believe that short simple questions and response options will be processed similarly by

[1] Dykema (2005) reports that between 1% and 20% of respondents "seek clarification" of a question. The category "seek clarification" includes asking for the question to be repeated along with other forms of clarifying requests, such as querying the meaning of a particular word. It is likely that only a minority of occurences of seeking clarification consist of asking for the question to be repeated. Similarly, Uhrig and Sala (2011) report that respondents "express uncertainty" in 7% of question administrations, a category which again includes asking for repetition along with other forms of uncertainty. Some requests for repetition may be made for reasons other than a failure to retain part of the question in memory. Other reasons would include a failure to hear (part of) the question, due to a distracting noise, multitasking (particularly in the case of telephone interviews), or an aural health problem (Lynn and Kaminska, 2012).

most respondents and that most or all of the relevant information will be correctly perceived. When the communicated information is greater in quantity or more complex in nature, differences will begin to emerge between respondents in the extent to which they fully perceive and process the information. Those with greater cognitive abilities should be more likely than others to correctly perceive and process all the relevant information and hence to correctly understand the question and the response task.

Visual perception varies considerably between individuals. This may partly be because individuals are able to assert their own preference regarding where to focus their attention and the order in which to attend to competing visual stimuli. It may also be because the processing of visual information is more cognitively demanding, so perception is more strongly associated with certain cognitive abilities than in the case of aural perception. Consequently, the benefits of visual stimuli in surveys may depend on the cognitive ability of the respondent.

In this paper we test most of the following general hypotheses about the effects of channels of communication (all other than H2a and H1c, which are included here only in order to present a complete framework). Each hypothesis is addressed by specific tests based on experiments described subsequently.

H1. Visual rather than aural presentation of a survey question or response options tends to increase the propensity that a respondent will correctly understand the question;

H2. Visual rather than aural presentation of a survey question or response options tends to reduce the propensity that a respondent will satisfice in responding;

H1a, H2a. Effects on understanding or satisficing will tend to be stronger when both question and response options are presented visually than when only the response options are presented visually, other things being equal;

H1b, H2b. Effects on understanding or satisficing will be mediated by the cognitive ability and motivation of the respondent. Specifically, positive effects of visual rather than aural presentation (better understanding; less satisficing) will be stronger for respondents of higher ability and motivation and weaker, absent or reversed for respondents of lower ability and motivation;

H1c, H2c. Effects on understanding or satisficing will be mediated by characteristics of the question. Specifically, positive effects of visual presentation will be confined to questions that are relatively difficult or burdensome and will be stronger for non-factual questions that factual ones.

3. Data and Methods

We address our hypotheses through analysis of four different experiments, each involving between one and six survey items. Each experiment addresses a number of our hypotheses, though no one experiment addresses all hypotheses. Some hypotheses are addressed by comparing experiments or comparing items.

The experiments formed part of a larger study in which around 1,800 respondents to the Omnibus Survey of the UK National Centre for Social Research were randomly allocated to a follow-up interview in one of three modes, CAPI, CATI or CAWI. The follow-up interviews were achieved with 1,138 persons aged 19 or over, representing participation rates of 73% for the CAPI sample, 69% for the CATI sample and 47% for the CAWI sample. The initial Omnibus Survey involved a random sample of the adult population (aged 18 or over) of Great Britain, and achieved a response rate (AAPOR RR1) of 64%. The net response rate for our three sub-samples could therefore be estimated to be around 47%, 44% and 30% respectively. We would note, however, that our inferences depend only on the random allocation conditional on response to the initial survey, not on the randomisation in the initial sample selection. Field work for the follow-up interviews on which our study is based was carried out in January to March 2009.

Our indicator of the extent of respondent understanding of a question (ordinal items only) is the observed correlation between the item and other theoretically-associated items.

We use two indicators of satisficing. These do not indicate satisficing behaviour for particular respondents. Rather, comparisons between subgroups of the distribution of the indicator indicate the relative extent of satisficing in the subgroups. For ordinal scales with a mid-point (all four experiments) we use as an indicator the proportion of respondents choosing the middle or extreme response categories, e.g. categories 1, 4 or 7 of a 7-point scale[2].

[2] This follows the approach of Kaminska, McCutcheon and Billiet (2010), who combine the middle-point indicator of satisficing used by Krosnick, Narayan and Smith (1996) with the extreme-point indicator used by Belli, Herzog and van Howeyk (1999).

Additionally, for fully-labelled items (experiments 1 and 2) we use indicators of the proportion of 'early' response categories chosen (primacy effect) and of the proportion of 'late' response categories chosen (recency effect)[3]. Our primacy and recency indicators are binary variables indicating choice of the first (or last) *n* response categories. The value of *n* differs between the items, depending on the number of response categories offered, their substantive interpretation, and the empirical distribution and for primacy is documented in Table 2.

Our indicators of cognitive ability and motivation are imperfect proxy measures but should be correlated with the respective dimension of interest. We use two binary indictors of cognitive ability: whether or not the respondent is aged over 65 (following Alwin and Krosnick 1991; see also Cohen 1987) and whether or not the respondent has any educational qualifications (see Ceci 1991). Our single binary indicator of motivation is based on an interviewer-coded item regarding whether they perceived that the respondent made an effort to answer questions to the best of their ability[4].

To test for effects of channel of communication on respondent understanding (H1) we compare between-item correlations. The true association between each ordinal item of interest and each related item can be assumed to be monotone, but not necessarily linear. We therefore use Spearman's rank correlation coefficient to measure association. Differences in correlation are tested by using the Fisher transformation of the Spearman rank correlation coefficient to derive z-scores as proposed by Choi (1977). We interpret lower correlation with the non-visual version to indicate the success of the visual version at improving understanding.

To test for effects of channel of communication on satisficing (H2) we fit logit models with the respective satisficing indicator as the dependent variable and treatment group as a predictor variable. For the first indicator (middle and extreme responses), we interpret a lower proportion to indicate less satisficing. For the primacy and recency indicators we interpret a difference between treatment groups as indicating that channel of communication affects satisficing, but we cannot directly infer the direction of the difference. To test H2b we assess interactions between treatment group and the indicators of cognitive ability and of

[3] Following Krosnick and Alwin (1987) and Schwarz, Hippler and Noelle-Neumann (1992).
[4] See section A2 of the appendix for wording. Though the item had five response options, 79% of CAPI respondents were coded to the first category, "always". Consequently, our binary indicator contrasts this response with the other four options.

motivation. To test H2c we compare effects between factual and non-factual questions and between difficult and less difficult questions.

In all models sex, age (banded), economic activity status, marital status, ethnic group and level of qualifications are included as controls. These variables were chosen primarily through identification of differences in sample composition between modes. To some extent, then, the models control for differential non-response between modes.

4. Experiment 1: visual vs aural communication of response options in face-to-face interviews

4.1 Experimental design

Show cards are commonly used in face-to-face surveys (Lynn 2004) for questions with five or more answer categories. This is done mainly because it is believed that visual display of response options makes the task of responding easier. The use of show cards standardises the visual presentation of information in an interview for which the primary channel of communication is aural. Experiment 1 investigates the effects of using show cards.

Face-to-face respondents were randomly allocated to two treatments. In the first, they were shown a card listing response options and were asked to select their answer from the card. In the second treatment, no cards were used: the response options were communicated solely by the interviewer reading them out, in the same order that they were listed on the card. Six survey items were treated in this way. Two of these were satisfaction questions (Q43 regarding household waste collection and recycling and Q44 street cleaning,) and four were factual (Q39 closest facility to home, Q75 housing type, Q81 monthly spend on leisure activities, Q82 length of residence in area). Two of the four factual items were deemed to be more difficult than the other two (Q39, Q81) as was one of the satisfaction questions (Q43). Question wordings and response options were identical in the two treatments. The only difference was that in the first treatment, the question was immediately followed by "*Please look at this card and tell me your answer*". Full question wordings are in section A1 of the appendix.

4.2 Analysis and Results

For one of the two 7-point satisfaction scales (Table 1) we observe significantly less satisficing with visual presentation, thus supporting H2. While we do not observe significant interactions with our indicators of cognitive ability or motivation, the main effect of visual presentation becomes non-significant when an interaction with age is added to the model. The predicted odds ratios suggest that the main effect is driven by respondents aged under 65, while the effect has the opposite sign for respondents 65 or over (predicted odds ratios for visual relative to aural are 0.62 for under-65s and 1.18 for those aged 65+). This is consistent with H2b.

We find some significant differences between treatments in the extent that respondents choose the first or last categories. The results for first categories are summarised in Table 2 and show significant effects for three items. Results for last categories are not shown but largely show the same, but opposite, patterns. Results are all in the direction that would indicate a recency effect with visual presentation and/or primacy effect with aural presentation, in other words opposite to the findings of Krosnick & Alwin (1987) and others. Holbrook et al (2007) show that recency effects with aural communication may only occur when interviewers always read to the end of the list of response options. If interviewers allow respondents to interrupt and select a response before they have heard all of them, primacy effects may instead dominate. This may be what we are observing here. We did not have control over whether the interviewers always read the full list of options or whether they allowed respondents to interrupt. In consequence, the dominant form of shortcutting in both treatments may be to select an early response option, but this tendency may be stronger in the aural treatment. This would be consistent with H2. Thus, visual presentation affects not only the tendency to select extreme categories (Table 1) but specifically the tendency to select early categories (Table 2). The impact of the effect will therefore depend on which response options are presented early.

Table 1: Effects of Visual Communication in CAPI on Choice of Middle and Extreme Categories

		Q43 Satisfaction re. waste collection			Q44 Satisfaction re street cleaning		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Visual *(omitted = aural)*	OR	0.88	0.99	1.15	0.54	0.62	0.49
	P	*0.65*	*0.99*	*0.68*	***0.05***	*0.18*	***0.05***
65+ *(omitted = age 18-64)*	OR		0.62			0.77	
	P		*0.39*			*0.67*	
Visual *65+	OR		1.80			1.90	
	P		*0.42*			*0.40*	
No quals *(omitted = quals)*	OR			0.55			1.54
	P			*0.27*			*0.44*
Visual *no quals	OR			3.28			0.64
	P			*0.10*			*0.48*

N=203 (110 visual, 93 aural); Estimates from logit models in which the dependent variable is an indicator of choice of categories 1, 4 or 7, versus 2, 3, 5, or 6; OR denotes odds ratio; *P* denotes P-value for the null hypothesis of OR=1.0; na indicates cannot be estimated due to collinearity; for both items, the interaction of visual with interviewer-reported respondent effort was tested in a fourth model, but in both cases there was no significant effect of the interaction.

Table 2: Effects of Visual Communication in CAPI on Choice of Early Categories

		Q39 Closest facility	Q43 Satisfaction re. waste collection	Q44 Satisfaction re street cleaning	Q75 Housing type	Q81 Leisure spend	Q82 Time in area
Categories		1,2	1,2	1	1	1	1,2,3
Visual *(omitted = aural)*	OR	1.20	1.14	0.39	1.10	0.41	0.67
	P	*0.64*	*0.67*	***0.04***	*0.78*	***0.01***	*0.43*
65+ *(omitted = age 18-64)*	OR	1.09	1.47	1.31	1.49	1.28	0.74
	P	*0.64*	*0.004*	*0.21*	*0.01*	*0.08*	*0.17*
Visual *65+	OR	1.68	0.91	1.23	1.34	3.35	na
	P	*0.57*	*0.89*	*0.79*	*0.65*	*0.07*	
No quals *(omitted = quals)*	OR	6.85	1.94	1.35	0.67	3.44	na
	P	*0.09*	*0.33*	*0.71*	*0.51*	***0.04***	
Visual *no quals	OR	0.09	1.20	1.63	0.18	0.55	na
	P	***0.05***	*0.84*	*0.62*	*0.09*	*0.44*	
Low effort *(omitted = high effort)*	OR	0.76	0.35	0.80	0.80	2.16	0.84
	P	*0.71*	*0.05*	*0.99*	*0.68*	*0.21*	*0.84*
Visual *low effort	OR	0.77	4.11	1.50	1.40	0.91	na
	P	*0.80*	*0.07*	*0.99*	*0.67*	*0.92*	

N=203 (110 visual, 93 aural); Odds ratio (OR) for main effect of visual is from a model with no interaction terms; OR for each interaction and its associated main effect is from a model with only one interaction; hence, each column represents four models; *P* is the P-value for the null hypothesis of OR=1.0; na indicates cannot be estimated due to collinearity.

We also find some evidence in support of H2b. The main effect for Q81 of less primacy with visual presentation is driven by respondents aged under 65. In fact, the effect goes in the opposite direction for respondents aged 65+, though it does not quite reach statistical significance. (Predicted relative odds of selecting the first category for visual versus aural presentation are 0.41 for under-65s and 1.37 for those aged 65+). Similarly, the main effect for Q39 is driven by respondents with no qualifications: there is no effect amongst those with qualifications (predicted relative odds of selecting the first category for visual versus aural presentation are 1.20 for qualified respondents and 0.11 for those with no qualifications). There is also an interaction between respondent motivation and visual presentation that is of borderline significance (P=0.07) for Q43, suggesting that visual presentation may increase primacy effects only for less motivated respondents. Predicted relative odds of selecting the first category for visual versus aural presentation are 1.1 for motivated respondents and 4.7 for unmotivated respondents.

We test H1 for Q81 (Q82 does not have any theoretically-associated items in the questionnaire, while the other two factual items are not ordinal). For this item, we expect the visual treatment to improve respondent understanding of the question-answering task by reducing confusion about the boundaries of response categories. Confusion would attenuate the correlation between responses to this item and other related survey items. We find some support for H1 (Table 3). Correlation of monthly spend on leisure activities with employment status is significant only with visual presentation. Correlation with gross income is significant with both treatments, but stronger with visual presentation.

Table 3: Correlations between Monthly Spend on Leisure (Q81) and Related Items

	Expected sign of correlation	$r(x_n,y)$	$r(x_v,y)$	*P*
Employment status (employed)	+	0.08	0.21	*
Gross income	+	0.27	0.30	*

r is the sample Spearman rank correlation coefficient; x_n is the non-visual version of the importance of money item; x_v is the visual version; *P* is the *P*-value relating to a null hypothesis of $R(x_n,y)- R(x_v,y)=0$, where $\{R\}$ are the respective population rank correlation coefficients. N=203 (110 visual, 93 non-visual; * indicates $0.05 > P > 0.01$; ** indicates $P < 0.01$

It is not possible to formally test H2c with the small set of survey items in this experiment, but we note that significant effects were found for both of the difficult factual questions and neither of the easy factual questions, a pattern which is at least consistent with H2c. However,

an effect was also found for the satisfaction question that was deemed to be the easier of the two.

5. Experiment 2: CAWI (visual) vs CATI (aural)

5.1 Experimental design

One possible criticism of the comparisons in the previous section is that the treatment involving show cards may not rely solely on visual communication. Interviewers may sometimes read out the response options, for example when a respondent appears to be struggling to read the options for themselves, perhaps due to poor literacy skills or poor eyesight. Some respondents may even explicitly ask the interviewer to read out the options for them. This is important as the extra channel of communication may be used precisely for the respondents who would otherwise be particularly susceptible to the response effect in which we are interested.

Such uncertainty about possible mixing of the channels of communication is avoided in other modes. In CAWI communication of the response options is purely via the visual channel. Similarly, in CATI we can be sure that communication is entirely aural. Therefore, we compare here the same items between CAWI and CATI respondents. It must be recognised, of course, that we are here substituting one source of possible confounding with another. In comparing CAWI with CATI the absence of an interviewer is no longer held constant. However, we believe that interviewer presence is unlikely to have had much influence on the data analysed here. Interviewer presence can influence survey responses through increased control over the question answering process or through elicitation of social norms. Regarding control, our CAWI interview already controls quite tightly certain aspects such as question order. Pace of the interview is perhaps the one feature that could differ between the treatments. Regarding social norms, none of the items that we consider here have particularly strong social desirability connotations nor are of the form that makes them susceptible to acquiescence. Interviewers can also motivate respondents, but any such effect would go in the opposite direction to the hypothesised effect of channel of communication (less satisficing with visual communication), so this will make our tests conservative. A further difference between experiments 1 and 2 is that the question stem was delivered aurally in both treatments in experiment 1; it was only for the response options that the channel of

communication differed. In experiment 2, the question stem was delivered via the same channel as the response options.

We consider here the same six items as in the previous section. Administration in CATI was identical to that described for the CAPI aural treatment. In CAWI, for each item the question wording and response options appeared on a single screen, with a radio button alongside each option. The words "Select one circle only" appeared immediately before the response options. An example of the layout appears in section A3 of the appendix.

5.2 Analysis and Results

We expect differences between CAWI and CATI to be similar to differences between the visual and aural treatments in experiment 1 if they are brought about primarily by the channel of communication of the response options. If differences are not similar, we would take this as an indication that the channel of communication of response options is not the dominant feature causing differences in measurement between treatments.

We observe no significant main effect of visual presentation on the extent of satisficing (Table 4), though there is a significant interaction with qualifications which suggests that CAWI increases the propensity to shortcut for respondents without qualifications but has no impact for qualified respondents. (Predicted relative odds of selecting extreme or middle options for visual relative to aural are 0.7 for respondents with qualifications and 4.0 for those without). Thus, we find some support for H2b.

For Q81, CAWI respondents were less likely than CATI respondents to select one of the last two categories (Table 5) and correspondingly more likely to select the first category (results not shown). For Q82 a significant interaction is observed between channel of communication and motivation. Predicted relative odds of selecting one of the last two categories for visual relative to aural are 0.99 for motivated respondents and 0.20 for less motivated respondents, again providing some support for H2b. The experiment provides no support for H2c as significant effects of channel of communication were found for one of the two difficult factual questions and one of the two less difficult ones.

Table 4: Effects of Visual Communication (CAWI vs CATI) on Choice of Middle and Extreme Categories

		Q43 Satisfaction re. waste collection			Q44 Satisfaction re street cleaning		
		Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Visual *(omitted = aural)*	OR	0.79	0.72	0.76	0.89	0.92	0.74
	P	*0.32*	*0.19*	*0.28*	*0.62*	*0.73*	*0.24*
65+ *(omitted = age 18-64)*	OR		1.43			0.53	
	P		*0.52*			*0.29*	
Visual *65+	OR		1.74			0.81	
	P		*0.47*			*0.80*	
No quals *(omitted = quals)*	OR			0.77			0.50
	P			*0.61*			*0.23*
Visual *no quals	OR			1.27			5.37
	P			*0.75*			***0.03***

N=319 (153 CAWI, 166 CATI); Estimates from logit models in which the dependent variable is an indicator of choice of categories 1, 4 or 7, versus 2, 3, 5, or 6; OR denotes odds ratio; *P* denotes P-value for the null hypothesis of OR=1.0; for both items, the interaction of visual with interviewer-reported respondent effort was tested in a fourth model, but in both cases there was no significant effect of the interaction.

Table 5: Effects of Visual Communication (CAWI vs CATI) on Choice of Late Categories

		Q39 Closest facility	Q43 Satisfaction re. waste collection	Q44 Satisfaction re street cleaning	Q75 Housing type	Q81 Leisure spend	Q82 Time in area
Categories		4,5,6,7	6,7	6,7	6	6,7	7
Visual *(omitted = aural)*	OR	1.20	0.85	0.56	2.48	0.25	0.99
	P	*0.53*	*0.61*	*0.16*	*0.14*	***0.002***	*0.97*
65+ *(omitted = age 18-64)*	OR	1.09	0.74	0.75	0.66	0.80	2.16
	P	*0.55*	*0.04*	*0.12*	*0.11*	*0.28*	*<0.001*
Visual*65+	OR	1.36	na	na	na	0.99	0.51
	P	*0.65*				*0.99*	*0.20*
No quals *(omitted = quals)*	OR	1.90	2.67	3.63	na	2.90	na
	P	*0.32*	*0.11*	*0.06*		*0.08*	
Visual *no quals	OR	0.64	na	0.94	na	0.17	na
	P	*0.61*		*0.95*		***0.06***	
Low effort *(omitted = high effort)*	OR	1.41	1.18	1.31	3.10	0.65	2.07
	P	*0.58*	*0.77*	*0.66*	*0.24*	*0.61*	*0.22*
Visual*low effort	OR	0.71	0.44	na	0.72	na	0.16
	P	*0.68*	*0.41*		*0.79*		***0.02***

N=319 (153 CAWI, 166 CATI); OR for main effect is the odds ratio for CAWI, relative to CATI, from a model with no interaction terms; OR for an interaction is the odds ratio for CAWI respondents with the relevant characteristic (aged 65 or over, with no educational qualifications, with low effort) from a model with only that interaction; *P* and na defined as for Table 1.

6. Experiment 3: Visual communication of end-labelled response scales in CAWI

6.1 Experimental design

Christian and Dillman (2004) demonstrate that with an end-labelled numeric scale, visual communication of the direction of the scale reduces respondent confusion about the direction and improves the criterion validity of the measures. However, the response scale that Christian and Dillman tested may be thought of as having counter-intuitive directionality as high values were associated with dissatisfaction and low values with satisfaction. This would appear to contravene the interpretive heuristics used by survey respondents (Tourangeau et al 2007) and could be expected to lead to respondent confusion. We implemented a similar experiment, but with a response scale with intuitive directionality (high values indicating high importance and low values indicating low importance).

CAWI respondents were randomly allocated to receive one of two versions of a question that asked how important money is to them, with response options on a scale of 1 to 10. One version of the question presented ten radio buttons, displayed linearly, representing the ten response categories, but with only the end categories labelled. The other version presented a box into which the respondent was asked to type the number corresponding with their response category. The question wording and end labels were identical in the two versions (see section A4 of the appendix).

6.2 Analysis and Results

We identify two ways in which improved understanding could be detected. First, we suggest that categories 2 to 9 will be chosen more frequently with the visual version as they are given more attention. This is because, though both question formats explicitly display the end categories, only in the visual version are the other eight categories represented visually. We therefore suppose that the visual version will increase the chance that a respondent understands that there are ten ordered response options. Second, following Christian and Dillman (2004), we suppose that the visual treatment improves respondent understanding of the question-answering task by reducing confusion about the direction of scale. Confusion about direction would attenuate the correlation between responses to this survey item and other related survey items. We therefore compare the two treatments in terms of correlations with related items.

The proportion of respondents choosing one of categories 2 to 9 is higher with the visual treatment (OR = 2.84; P=0.02, based on logit model with treatment indicator and control variables as predictors, results not shown). This supports H1. Interactions of treatment with age, qualifications and effort are not significant (P>0.05), so we find no support here for H1b.

We identify nine items that we would expect to be correlated with the importance of money item. These are listed in Table 6, along with the Spearman rank correlation coefficient for each question format and the P-value for the test of difference between them. For six of the nine items, there is a significant difference in correlation ($P < 0.05$) between the question formats. For five of these six, the direction of the difference indicates greater validity with the visual treatment. Specifically, there are three items for which the observed correlation is in the expected direction with the visual treatment and in the opposite direction with the non-visual treatment. For one item (Q3), the correlation is in the expected direction with both treatments, but is much stronger with the visual treatment. For one item (Q64) the correlation is in the unexpected direction with the non-visual treatment and is close to zero with the visual treatment. There is, however, one item (Q66) for which the correlation is in the opposite direction with the visual treatment and is close to zero with the non-visual treatment. These findings support H1.

For each of the five items for which we observed greater validity with the visual treatment, we examined whether this effect interacted with either cognitive ability or motivation. We compared differences in the correlations between sample subgroups (Table 7). The findings provide little evidence that effects differ between high and low cognitive ability groups or between more or less motivated respondents, and thus no support for H1b. Almost all of the effects tested (28 out of 30) are estimated to be in the direction of greater validity with visual presentation and the majority of these are statistically significant. For Q3 we see that the increase in correlation is significant only for the less able and less motivated respondents, but for all other questions there is no such pattern. (We note that the statistical power is considerably less for the less able and less motivated respondents as these are much smaller groups.)

Visual presentation reduces the extent of short-cutting (Table 8), consistent with H2. We do not find significant support for H2b, though the main effect of treatment becomes non-significant when interactions with either qualifications or motivation are added to the model.

Table 6: Correlations between Importance of Money Item and Related Items

	Expected sign of correlation	$r(x_n,y)$	$r(x_v,y)$	*P*
Q1 How well managing financially	+	0.11	0.02	-
Q3 Financial expectation a year from now	-	-0.06	-0.16	*
Q6 Monthly spend on eating out	-	0.11	0.09	-
Q64 Would rarely read small print before a financial decision	+	-0.23	-0.01	**
Q65 Would do a lot of research before a financial decision	-	0.01	-0.01	-
Q66 Would rarely talk to a financial advisor before a financial decision	+	0.02	-0.12	*
Q67 Would definitely talk to family/friends before a financial decision	-	0.09	-0.10	*
Q75 Housing type	+	-0.21	0.17	**
Q81 Monthly spend on leisure activities	-	0.22	-0.09	**

r is the sample Spearman rank correlation coefficient; x_n is the non-visual version of the importance of money item; x_v is the visual version; *P* is the *P*-value relating to a null hypothesis of $R(x_n,y)- R(x_v,y)=0$, where {*R*} are the respective population rank correlation coefficients. *N*=349 (166 visual, 183 non-visual; * indicates $0.05 > P > 0.01$; ** indicates $P < 0.01$

Table 7: Effect of Visual Presentation on Correlation between Related Items

		Under 65	65+	Quals	No quals	Effort	Low effort
Q3 Financial expectation a year from now (-)	Δ *r*	-0.06	-0.57	-0.08	-0.55	-0.06	-0.34
	P	*0.10*	***0.02***	*0.08*	***0.01***	*0.23*	***0.02***
Q64 Would rarely read small print before a financial decision (+)	Δ *r*	0.20	0.20	0.26	-0.42	0.19	0.31
	P	***0.002***	*0.09*	***0.000***	***0.05***	***0.002***	***0.03***
Q67 Would definitely talk to family/friends before a financial decision (-)	Δ *r*	-0.27	0.31	-0.18	-0.47	-0.19	-0.10
	P	***0.004***	***0.05***	***0.01***	***0.004***	***0.01***	*0.23*
Q75 Housing type (+)	Δ *r*	0.39	0.09	0.37	0.41	0.30	0.77
	P	***0.002***	*0.09*	***0.003***	***0.02***	***0.002***	***0.002***
Q81 Monthly spend on leisure activities (-)	Δ *r*	-0.30	-0.27	-0.27	-0.47	-0.27	-0.28
	P	***0.002***	*0.06*	***0.005***	***0.02***	***0.002***	***0.03***

Δ *r* is the difference in the Spearman rank correlation coefficient between the visual and non-visual treatments; *P* and *N* both as in Table 6.

Table 8: Effects of Visual Communication* on Choice of Middle and Extreme Categories

		Model 1	Model 2	Model 3	Model 4
Visual	OR	0.59	0.60	0.66	0.62
(omitted = aural)	*P*	***0.03***	***0.05***	*0.11*	*0.07*
65+	OR		0.76		
(omitted = age 18-64)	*P*		*0.64*		
Visual *65+	OR		1.48		
	P		*0.61*		
No quals	OR			6.72	
(omitted = quals)	*P*			*0.12*	
Visual *no quals	OR			0.37	
	P			*0.22*	
Low effort	OR				1.61
(omitted = high effort)	*P*				*0.65*
Visual*low effort	OR				0.71
	P				*0.63*

Estimates from logit models in which the dependent variable is an indicator of choice of categories 1, 5 or 10; OR denotes odds ratio; *P* denotes P-value for the null hypothesis of OR=1.0. *N* as in Table 6

7. Experiment 4: CAPI with show cards (visual) vs CATI (aural)

7.1 Experimental design

Experiment 3 compared end-labelled numeric scales with and without visual display of the scale, in the context of self-completion without interviewer involvement. Experiment 4 investigates the same effect, but in a context where the question stems are delivered aurally by an interviewer rather than visually. The experiment compares four end-labelled 7-point numeric scales between CAPI, with the scale displayed on a show card, and CATI. In this experiment the presence or absence of visual display of the response scale is of course confounded with the proximity of the interviewer (physically present or on the telephone), so we cannot rule out that differences in social desirability effects could contribute to any observed differences in response. However, we do not believe that any of the four items have strong social desirability connotations, so we expect any such effects to be small. The questions – two satisfaction items and two factual items – are reproduced in section A4 of annex A.

7.2 Analysis and Results

A significant difference is observed in the marginal distribution of responses for two of the four items, and in the proportion choosing categories 2 to 6 for one of the items (Table 9). As hypothesised, the proportion is higher in CAPI. In logit models (results not presented), no significant interactions of treatment with the indicators of cognitive ability or motivation were found for Q16, Q17 or Q74. There were interactions of borderline significance for Q68 with age (P=0.09) and qualifications (P=0.07). We therefore find mild support for H1 but little or no support for H1b.

There is a difference in the proportion choosing category 4, the mid-point, for three of the items (Table 9). In all three cases the proportion is larger in CAPI and the differences are substantial. We suspect that respondents find it easier to identify the mid-point when it is presented visually.

As regards correlations with other items, this can only be tested for Q74, as the other items do not have other natural correlates amongst the questionnaire items. We identify two items that could be expected to correlate with Q74, the frequency of purchase of teas and coffees outside the home. These items are employment status and income. We find that each of these correlations are positive in CATI, as expected, but are not significantly different from zero in CAPI. This does not therefore support H1.

Table 9: Effect of Visual Presentation of Response Scale on Response Distribution

	Distribution	Categories 2-6			Category 4 (mid-point)		
	P	CAPI %	CATI %	P	CAPI %	CATI %	P
Q16 Satisfaction with democracy	0.004 ($\chi^2(6)=19.0$)	91.0	80.7	0.004 ($\chi^2(1)=8.1$)	34.5	18.4	0.000 ($\chi^2(1)=12.9$)
Q17 Satisfaction with economy	0.93 ($\chi^2(6)=1.8$)	60.4	60.3	0.97 ($\chi^2(1)=0.0$)	9.0	7.2	0.50 ($\chi^2(1)=0.5$)
Q68 frequency of grocery shopping	0.000 ($\chi^2(6)=32.0$)	89.8	86.1	0.26 ($\chi^2(1)=1.3$)	29.9	13.0	0.000 ($\chi^2(1)=16.7$)
Q74 Purchases of teas and coffees	0.14 ($\chi^2(6)=9.6$)	84.2	85.2	0.77 ($\chi^2(1)=.08$)	10.2	4.3	0.02 ($\chi^2(1)=5.1$)

For all four items, there was less satisficing in the visual treatment (Table 10 and Table 11), though the difference only reached significance for Q68 (P=0.004) and was borderline for Q74 (P=0.08). For Q74, an interaction with qualification level was observed (predicted odds ratios for visual relative to aural are 0.52 for qualified people and 10.86 for unqualified people). These findings provide weak support for H2 and H2b.

Table 10: Effects of Visual Presentation of Response Scale on Choice of Middle and Extreme Categories: Satisfaction items

		Q16 Satisfaction with democracy				Q17 Satisfaction with economy			
		Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Visual *(omitted = aural)*	OR	0.83	0.67	0.72	0.89	0.88	0.84	0.88	0.85
	P	*0.39*	*0.08*	*0.18*	*0.63*	*0.54*	*0.42*	*0.59*	*0.50*
65+ *(omitted = age 18-64)*	OR		1.13				1.18		
	P		*0.22*				*0.08*		
Visual *65+	OR		3.39				1.34		
	P		***0.01***				*0.52*		
No quals *(omitted = quals)*	OR			1.09				1.48	
	P			*0.81*				*0.29*	
Visual *no quals	OR			2.21				1.14	
	P			*0.12*				*0.79*	
Low effort (omitted = high effort)	OR				1.88				0.86
	P				*0.14*				*0.73*
Visual *low effort	OR				0.66				1.17
	P				*0.48*				*0.79*

N=384 (177 CAPI, 207 CATI); Estimates from logit models in which the dependent variable is an indicator of choice of categories 1, 4 or 7, versus 2, 3, 5, or 6; OR denotes odds ratio; *P* denotes P-value for the null hypothesis of OR=1.0.

Table 11: Effects of Visual Presentation of Response Scale on Choice of Middle and Extreme Categories: Behavioural items

		Q68 frequency of grocery shopping				Q74 Purchases of teas and coffees			
		Model 1	Model 2	Model 3	Model 4	Model 1	Model 2	Model 3	Model 4
Visual *(omitted = aural)*	OR	0.53	0.55	0.51	0.56	0.65	0.62	0.52	0.71
	P	***0.004***	***0.01***	*0.01*	*0.02*	*0.08*	*0.08*	*0.02*	*0.20*
65+ *(omitted = age 18-64)*	OR		1.02				1.07		
	P		*0.85*				*0.54*		
Visual *65+	OR		0.81				1.32		
	P		*0.66*				*0.61*		
No quals *(omitted = quals)*	OR			1.09				0.37	
	P			*0.82*				*0.04*	
Visual *no quals	OR			1.22				3.13	
	P			*0.70*				*0.08*	
Low effort *(omitted = high effort)*	OR				1.84				1.08
	P				*0.14*				*0.87*
Visual *low effort	OR				0.69				0.49
	P				*0.54*				*0.37*

See note to Table 10.

8. Summary and Discussion

The hypothesis that visual presentation increases respondent understanding (H1) was addressed by two experiments (CAWI with visual presentation of response options *vs*. CAWI with write-in of response; CAPI with show cards *vs*. CATI). Both experiments supported the hypothesis, though neither provided evidence in support of the hypothesis (H1b) that this effect would be stronger for respondents with greater cognitive ability or greater motivation. We have only a weak test of the hypothesis (H1a) that effects on understanding will be greater when both the question and response options are presented visually than when only the response options are presented visually. If H1a were true, we should expect stronger effects from experiment 3 than from experiment 4. The findings are consistent with this. With experiment 3, H1 found support from the test of use of all scale points and from five of nine tests of validity. With experiment 4, support was provided by only one of four tests of use of all scale points and from neither of two tests of validity.

We conclude that visual presentation of response options can indeed impact on respondent understanding. However, our experiments addressed only two specific aspects of

understanding in the context of end-labelled ordinal scales. There are many other aspects of understanding that could be affected by visual presentation and further research is needed to provide a more complete picture of how and when visual presentation improves understanding.

All four of our experiments provide some evidence that visual presentation produces less satisficing (H2), though the strength and nature of the evidence varies. Three of the four experiments additionally support the hypothesis (H2b) that this effect is dependent on the cognitive ability or motivation of the respondent. The hypothesis (H2c) that reduced satisficing will be restricted to difficult questions and will be more likely with non-factual questions is addressed by only two experiments. Our findings are inconsistent, but the test is weak. Our similarly weak test of hypothesis H2a - that effects on satisficing will be greater when both the question and response options are presented visually - is that effects should be stronger from experiment 2 than from experiment 1 and from experiment 3 than from experiment 4. We find the latter but not the former.

We conclude that visual presentation can reduce the extent of shortcutting, though the effect may be restricted to respondents of higher cognitive ability and greater motivation. We believe that this effect is more likely with difficult questions, though our evidence for this is weak. We would note that in our experiments visual presentation was employed as an alternative either to aural presentation (experiments 1 and 2) or to an absence of explicit presentation (experiments 3 and 4). Our hypotheses would suggest that it could be useful to consider the possibility, in certain circumstances, of using visual presentation as an addition to aural presentation, rather than as a replacement for it. While we have shown that (some) respondents with higher cognitive ability appear to gain greater understanding from visual presentation than from aural presentation, it seems plausible to us that the opposite could be true for (some) lower ability respondents. If true, this could suggest that using both channels in combination may be a useful option for maximising understanding amongst all respondents. This would rely on respondents paying attention to the channel that is most useful to them, which in turn raises design and implementation questions to ensure that this happens. And whether this is a worthwhile strategy is likely to depend on the nature of the likely misunderstanding of a question as well as the heterogeneity of the sample in terms of cognitive ability. We think that the issues associated with dual-channel communication of survey questions and response options warrant investigation.

Three of our experiments involved comparing channels of communication of response options while holding constant the channel of communication of the question stem. In two cases the stem was presented aurally and in the third, visually. Experiment 2, however, simultaneously varied the channel of communication of both the stem and the response options. The belief that considerations regarding the best channel of communication could differ between stems and response options is ingrained in the common use of show cards during face-to-face interviews. We believe that practice in this respect is based on little theory or evidence, however. Further research should attempt to separate out the effects of channel of communication of the stem and of the response options and to identify the best combinations of channels for the three main pieces of information communicated when administering a question: the question stem, the response options, and the answer.

Regarding show cards, little is known about how they are actually used in practice or how different ways of using them affect measurement. In particular, we do not know how frequently the card is shown to a respondent while the interviewer is reading out the question, as opposed to only after she has finished reading it out. And if the card is shown simultaneously to reading the question, we do not know to which channel of communication the respondent pays more attention. Without such knowledge, it is hard to know how we should ideally like cards to be used, from a measurement error perspective.

We should also note that our measures of cognitive ability and motivation represent a limitation of our study. Future studies of the moderating effects of these characteristics on measurement error would benefit from the development of better measures that can be easily implemented in a survey setting. Good measures of eyesight and hearing ability would also be helpful.

References

Alwin, Duane F. and Jon A. Krosnick. 1991. "The Reliability of Survey Attitude Measurement: The Influence of Question and Respondent Attributes." *Sociological Methods and Research* 20(1):139-181.

Belli, Robert F., A. Regula Herzog and John Van Howeyk. 1999. "Scale Simplification of Expectations for Survival: Cognitive Ability and the Quality of Survey Responses." *Cognitive Technology* 4:29-38.

Ceci, Stephen J. 1991. “How Much Does Schooling Influence General Intelligence and Its Cognitive Components? A Reassessment of the Evidence.” *Developmental Psychology* 27:703–22.

Choi, S.C. 1977. “Test of Equality of Dependent Correlations.” *Biometrika,* 64 (3):645–647.

Christian, Leah M. & Dillman, Don A. 2004. “The influence of graphical and symbolic language manipulations on responses to self-administered questions.” *Public Opinion Quarterly* 68:57-80.

Cohen, G. 1987. “Speech Comprehension by The Elderly: The Effects of Cognitive Change.” *British Journal of Audiology* 21:221-226.

Dickens, W.T. 2008. “Cognitive Ability”, in S.N. Durlauf and L.E. Blume (ed.s), *The New Palgrave Dictionary of Economics*, 2nd Edition, Palgrave Macmillan.

Dillman, Don A., Smyth, Jolene D. and Christian, Leah M. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Hoboken, New Jersey: Wiley.

Duncan, Starkey Jr. 1969. “Nonverbal Communication.” *Psychological Bulletin* 72(2):118-137.

Dykema, Jennifer. 2005. “An Investigation of the Impact of Departures from Standardized Interviewing on Response Errors in Self-Reports about Child Support and Other Family-Related Variables.” Paper Presented at the *Annual Conference of the American Association for Public Opinion Research*, Miami Beach, Florida.

Holbrook, Allyson L., Jon A. Krosnick, David Moore and Roger Tourangeau. 2007. “Response Order Effects in Dichotomous Categorical Questions Presented Orally.” *Public Opinion Quarterly*, 71:325-348.

Jobard, Gael, Mathieu Vigneau, Bernard Mazoyer, and Nathalie Tzourio-Mazoyer. 2007. “Impact of Modality and Linguistic Complexity During Reading and Listening Tasks.” *NeuroImage* 34:784-800.

Just, Marcel A., Patricia A. Carpenter, Timothy A.Keller, William F. Eddy, and Keith R. Thulborn. 1996. “Brain Activation Modulated By Sentence Comprehension.” *Science* 274:114-116.

Kaminska, Olena, Allan McCutcheon, and Jaak Billiet. 2010. “Satisficing Among Reluctant Respondents in a Cross-National Context.” *Public Opinion Quarterly* 74:956-984.

Kellogg, Ronald T. 2007. *Fundamentals of Cognitive Psychology*. Los Angeles: SAGE Publications.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5:213-236.

Krosnick, Jon A. and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51:201-219.

Krosnick, Jon A., Narayan, S. and Smith, W.R. 1996. "Satisficing in Surveys: Initial Evidence." In Braverman, M. T. and J. K. Slater (Eds.), *Advances in Survey Research* 29-44. San Francisco: Jossey-Bass.

Lynn, Peter. 2004. "Show card", in *Encyclopedia of Research Methods for the Social Sciences* (eds. A. Bryman, M. Lewis-Beck and T.F. Liao). London: Sage Publications.

Lynn, Peter and Olena Kaminska. 2012. "The Impact of Mobile Phones on Survey Measurement Error." *Public Opinion Quarterly*, forthcoming, published online.

Payne, Stanley L. 1951. *The Art of Asking Questions*, Princeton NJ: Princeton University Press.

Schwarz, Norbert, Fritz Strack, Hans-Jürgen Hippler, and George Bishop. 1991. "The impact of administration mode on response effects in survey measurement." *Applied Cognitive Psychology* 5: 193-212.

Schwarz, Norbert, Hans-Jürgen Hippler and Elisabeth Noelle-Neumann. 1992. "A Cognitive Model of Response-Order Effects in Survey Measurement". In Norbert Schwarz and Seymour Sudman (ed.s) *Context Effects in Social and Psychological Research*, 187-199. New York: Springer Verlag.

Stern, Michael J., Don A. Dillman, and Jolene D. Smyth. 2007. "Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey." *Survey Research Methods* 1:121-138.

Sudman, Seymour and Norman M. Bradburn. 1974. *Response Effects in Surveys: A Review and Synthesis*. Chicago: Aldine.

Tourangeau, Roger, Mick P. Couper, and Frederick Conrad. 2007. "Color, Labels, and Interpretive Heuristics for Response Scales." *Public Opinion Quarterly* 71(1):91-112.

Uhrig, S.C. Noah and Emanuela Sala. 2011. “When Change Matters: An Analysis of Survey Interaction in Dependent Interviewing on the British Household Panel Survey.” *Sociological Methods and Research* 40(2):333-366.

Appendix: Question Wording

A1 Experiment 1 items

For each item, in treatment 1 the interviewer read out the response options, and in treatment 2 the interviewer handed the respondent a card listing the response options and read the words "Please look at this card and tell me your answer."

Q39. Which of the following is closest to where you live?

A primary school.....	1
A secondary school.....	2
A 6th form college.....	3
A river.....	4
A lake.....	5
A cinema.....	6
Or, a theatre.....	7

Q43. I would like you to tell me how satisfied or dissatisfied you are with local household waste collection, recycling collection and other recycling collection points.

Very satisfied.....	1
Moderately satisfied.....	2
Slightly satisfied.....	3
Neither satisfied nor dissatisfied.	4
Slightly dissatisfied.....	5
Moderately dissatisfied.....	6
Or very dissatisfied?.....	7

Q44. And how satisfied or dissatisfied are you with street cleaning?

Very satisfied.....	1
Moderately satisfied.....	2
Slightly satisfied.....	3
Neither satisfied nor dissatisfied.	4
Slightly dissatisfied.....	5
Moderately dissatisfied.....	6
Or very dissatisfied?.....	7

Q75. Which of these best describes your home?

Detached house.....	1
Semi-detached house	2
Terraced house.....	3
Bungalow	4
Flat in a block of flats.....	5
Flat in a house.....	6
Maisonette.....	7
Or other?	8

Q81. How much do you personally spend in an average month on leisure activities, and entertainment and hobbies, other than eating out?

Less than £20.....	1
£20 - £39.....	2
£40 - £59	3
£60 - £79	4
£80 - £99.....	5
£100 - £119.....…….	6
£120 - £139 .....…….	7
Or £140 or more.?.....	8

Q82. How long have you lived in this area?

Less than 12 months……..	1
12 months or more but less than 2 years…	2
2 years or more but less than 3 years…….	3
3 years or more but less than 5 years…….	4
5 years or more but less than 10 years……	5
10 years or more but less than 20 years..…	6
Or 20 years or longer……..	7

A2 Indicator of respondent motivation

INTERVIEWER: Did you feel that the respondent made an effort to answer the questions to the best of his or her ability?

Always…….. 1

Nearly always…….. 2

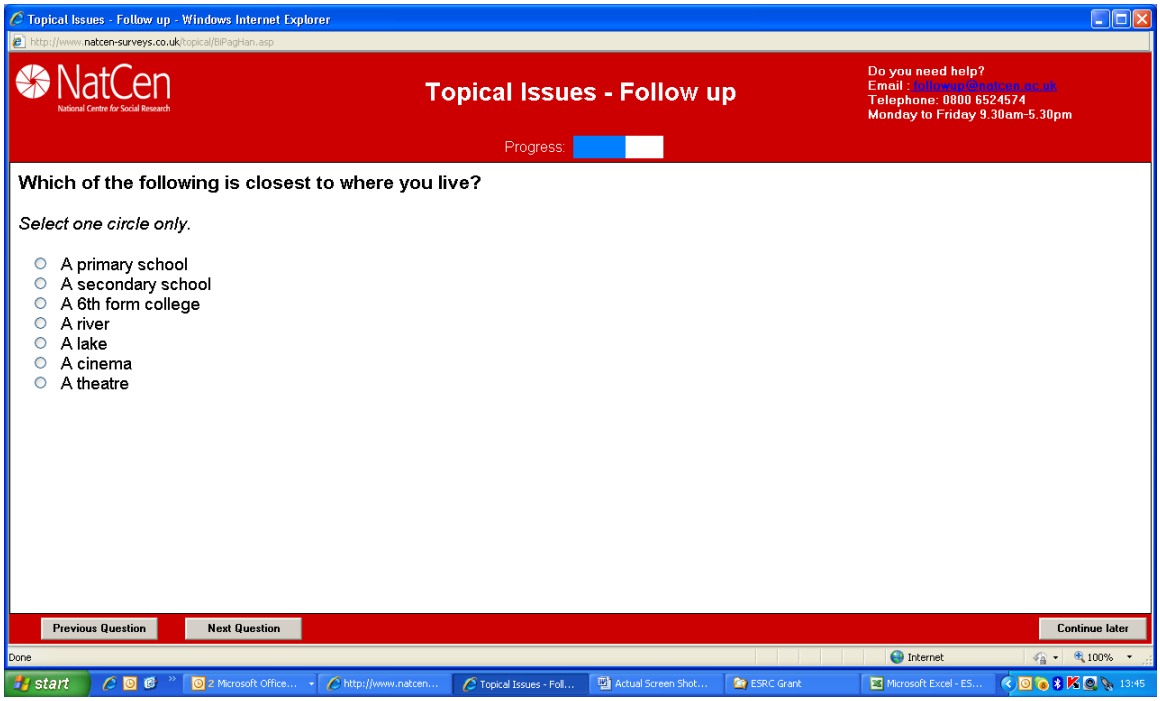
Most of the time ……..... 3

Some of the time ……4

Rarely ……5

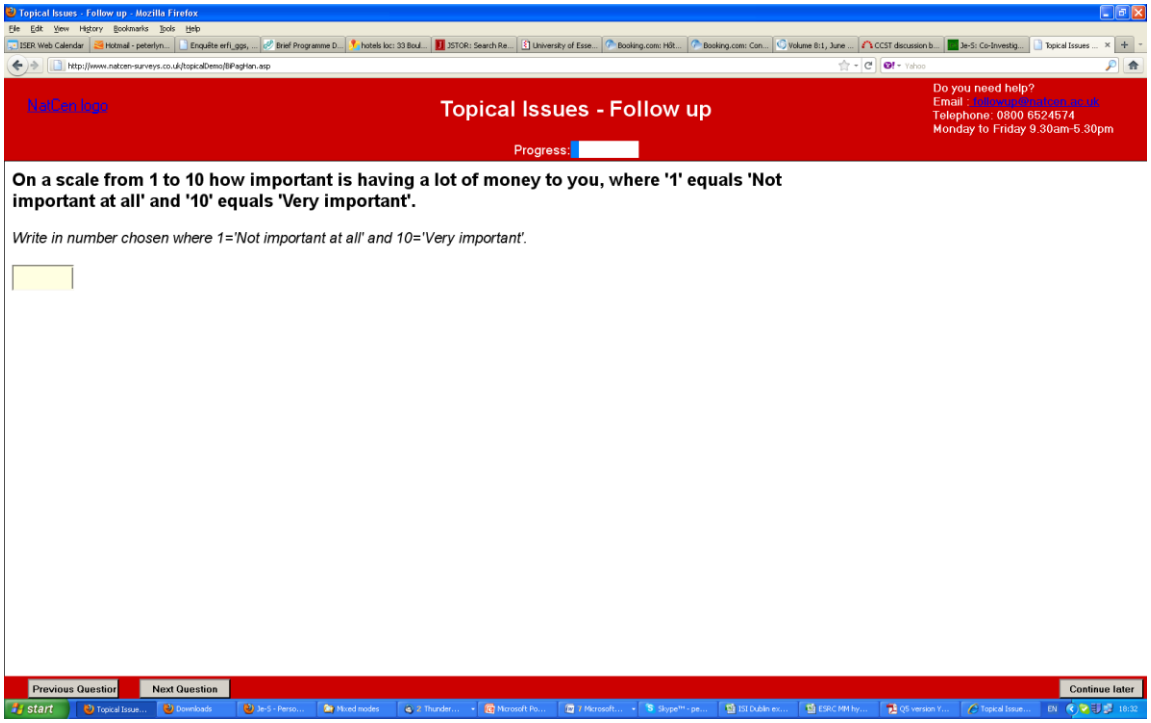
Codes 2 to 5 classified as "low motivation"

A3 Example of screen layout for CAWI implementation of fully-labelled categorical response options (Experiment 2)

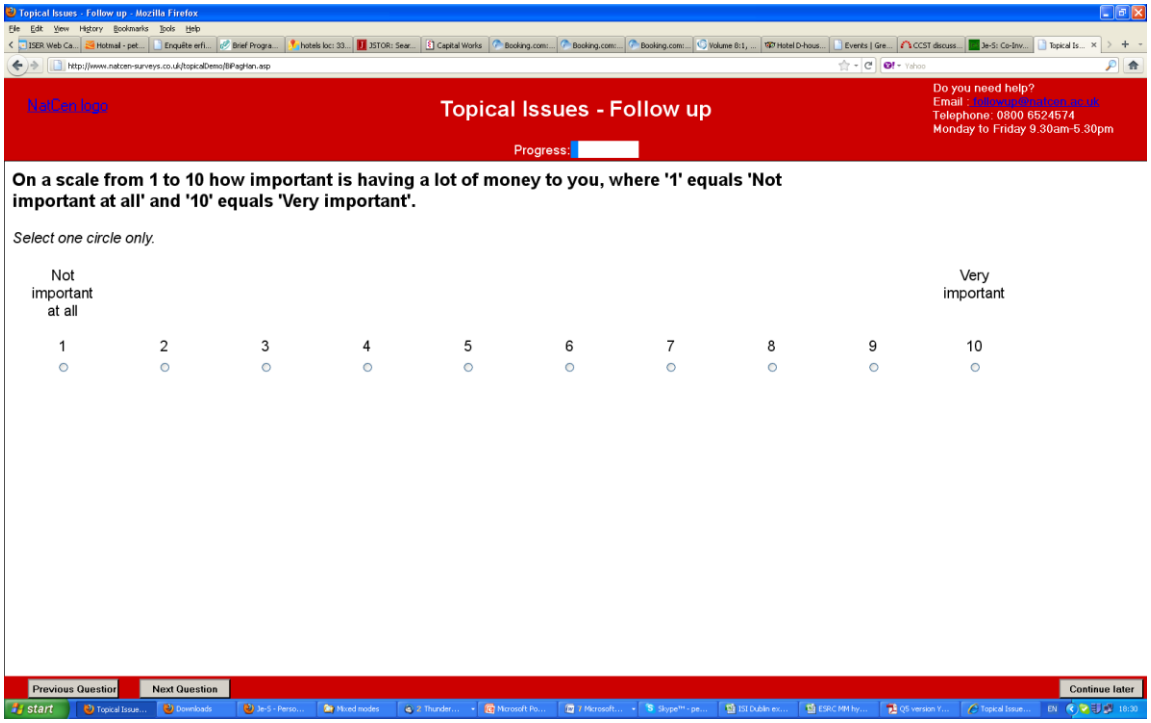


A4 Experiment 3 items

'Non-visual' version



'Visual' version



A5 Experiment 4 items

Visual version

GB16x. On the whole, how satisfied are you with the way democracy and personal freedom work in Great Britain? Please look at this card and tell me your answer.

Very Satisfied 1
2
3
4
5
6
Very dissatisfied…....7

GB17x. And on the whole, how satisfied are you with the present state of the economy in Great Britain? Please look at this card and tell me your answer.

Very Satisfied 1
2
3
4
5
6
Very dissatisfied…....7

FM68x.
The next question is about grocery shopping which includes food, drinks, cleaning products, toiletries and household goods. How often do you personally do grocery shopping? Please look at this card and tell me your answer.

Every day..... 1
2
3
4
5
6
Never ……..... 7

FM74x. In the last two weeks, how many teas, coffees and other hot beverages have you purchased outside the home? Please look at this card and tell me your answer.

None…..... 0
1
2
3
4
5
More than 25..... 6

Aural version

GB16x. On the whole, how satisfied are you with the way democracy and personal freedom work in Great Britain, where 1 is very satisfied and 7 is very dissatisfied? Please give me a number between 1 and 7.

GB17x. And on the whole, how satisfied are you with the present state of the economy in Great Britain, where 1 is very satisfied and 7 is very dissatisfied? Please give me a number between 1 and 7.

FM68x. The next question is about grocery shopping which includes food, drinks, cleaning products, toiletries and household goods. How often do you personally do grocery shopping, where 1 is every day and 7 is never? Please give me a number between 1 and 7.

FM74x. In the last two weeks, how many teas, coffees and other hot beverages have you purchased outside the home, where 1 is more than 25 and 7 is none? Please give me a number between 1 and 7.