Routledge
Taylor & Francis Group

OPEN ACCESS

Are there differences in responses to social identity questions in face-to-face versus telephone interviews? Results of an experiment on a longitudinal survey

Alita Nandi[a] and Lucinda Platt[b]

[a]Institute for Social and Economic Research (ISER), University of Essex, Colchester, UK; [b]Department of Social Policy and CASE, London School of Economics and Political Science, London, UK

ABSTRACT

This paper investigates the effect of interview mode (telephone vs. face-to-face) on responses to a 13-item module of identity questions covering distinct domains. With increasing moves towards mixed-mode implementation, especially in longitudinal surveys, establishing whether mode effects are likely to influence findings is of practical value. A growing number of studies explore mode effects; but the potential impact of mode on identity questions has not been investigated, even though such questions are increasingly being asked in multi-topic surveys. Adjusting for selection, we find little evidence for specific mode effects. The exception is responses on political identity: telephone responders are eight percentage points more likely to consider politics important to their identity. We do not find differences in data quality as measured by item non-response, straightlining, primacy and recency effects across modes. We conclude that mode effects are small for identity questions.

Introduction

In a context in which national surveys face declining response and increasing costs, sequential mixed mode designs have offered one potential route to maintaining sample sizes at reduced costs (Couper, 2011; de Leeuw, 2005; Dillman et al., 2009). This has been a particular consideration for longitudinal surveys where the establishment of panels has successfully been used as the basis for more economical modes of follow-up without pre-supposing loss of data quality. For example, the well-established US Panel Study of Income Dynamics now uses telephone to re-contact its respondents, the Longitudinal Study of Young People in England moved to mixed mode for its later sweeps, the National Child Development Survey introduced sequential mixed mode for the age 55 survey, while the UK Labour Force Survey involves quarterly telephone follow-ups to an initial face-to-face contact. However, such mixing of interview modes may come at the cost of non-comparability, particularly in relation to questions that are more sensitive to the mode of administration (Tourangeau & Smith, 1996).

CONTACT Alita Nandi ✉ anandi@essex.ac.uk

A number of studies have considered the potential impact of mixed mode designs on quality and accuracy of response, through investigation of 'mode effects' (see e.g. de Leeuw, 2005; Dillman, 2000; Dillman & Tarnai, 1988; Holbrook, Green, & Krosnick, 2003; Jäckle, Roberts, & Lynn, 2010; Lynn, 1998). As well as identifying those questions that tend to be most sensitive to mode effects, such as sensitive behavioural questions or attitudinal questions, the literature has also indicated ways in which the impact of mode can be mitigated by effective design (Dillman, 2000). While the differences in mode effects across telephone and face-to-face surveys tend to be rather small (de Leeuw, 2005), the direction of effects in terms of data quality have not always been shown to be consistent, with some suggesting higher quality in face-to-face (de Leeuw, 2005) and others in telephone surveys (Biemer, 2001; Groves, 1990). The bulk of investigation of mode effects has, however, focused on cross-sectional experimental and non-experimental contexts (de Leeuw, 2005), rather than testing for mode differences among those who are already respondents to an existing survey, as in the case of longitudinal surveys.

Capitalising on a unique experimental survey methodological resource, we test for mode effects on both data quality (as measured by item non-response, straightlining, primacy and recency effects) and on substantive responses to social identity questions, for a longitudinal sample. By social identity we refer to individuals' recognition of belonging to a particular social group, and their evaluation of the affective importance of such belonging (Tajfel, 1981; see also Hogg, 2006). Our suite of 13 identity questions do not represent a single, coherent, latent identity factor, but measure distinct aspects of a person's self-concept. Such measures are typically used by social psychologists to investigate identity in specific populations (e.g. Phinney, 1990; Verkuyten, 2007). More recently, analysis of identity in nationally representative samples is being eagerly taken up by a range of disciplines, utilising available measures in large-scale multi-topic surveys (e.g. Casey & Dustmann, 2010; Manning & Roy, 2010; Platt, 2014; Tilley, 2003). However, we know very little about how identity questions perform across modes. Evidence on the robustness of identity questions to a mixed-mode context could have practical implications for the analysis of identity across different surveys and the estimation of identity change over time.

We use an experiment involving random allocation to telephone or face-to-face mode and implement a 'unimode' design (Dillman, 2000) intended to remove sources of variation between the modes as far as possible. By these means, we argue, we should be able to isolate and evaluate any direct mode effects, i.e. those related solely to the physical presence or absence of the interviewer. This enables us to test competing expectations of mode effects on the varied domains of self-concept covered by our 13 identity questions.

By restricting our analysis to those who are existing respondents to a longitudinal study, we reduce the traditional advantage of face-to-face studies in terms of engaging respondents and establishing trust. Typically, face-to-face interviews have an advantage over telephone mode in terms of data quality due to the relative ease of establishing trust through rapport. However, this higher rapport can also increase the probability of socially desirable answers to sensitive questions (Krumpal, 2013). In longitudinal surveys, existing respondents to the survey, who respond at a second or later sweep (by whatever mode) can be assumed to be equally engaged with and trusting of the survey.

By these means, we were able to identify whether there were any genuine mode effects in the quality or nature of responses. We found few differences across the two modes. We did not find any differences in data quality as measured by straightlining, primary and recency effects but found small mode effects in item non-response for occupational identity, which could be fully accounted for by slight differences in the composition of face-to-face and telephone samples. In terms of response patterns, we found significant mode effects of eight percentage points in responses relating to political identity, with stronger assertions of political identity among telephone responders. This difference suggested that, under conditions of equal trust in the survey, telephone interviews may offer greater social distance and lead to more open responses in a domain (politics) that is recognised as being subject to substantial social desirability bias. However, we cannot dismiss the possibility that with a suite of 13 questions, one significant finding may be due to chance. On balance and under effective design conditions, identity questions appear robust to mode.

Background: existing literature and implications

Size of mode effects

The use of different modes for two consecutive interview sweeps or waves in longitudinal studies is well-established across a number of large-scale surveys, such as in the quarterly phone follow-ups to the UK Labour Force Survey, or the alternate phone and face-to-face interviews in the Health and Retirement Study. Moreover, there is long-standing use of phone in established panels such as the Panel Study of Income Dynamics. In a context in which there are increasing pressures to reduce survey costs on high quality surveys, sequential mixed mode designs within a sweep have offered an additional potential route to maintaining sample sizes at reduced costs (Couper, 2011; de Leeuw, 2005; Dillman, 2000). While, there is not yet consensus as to whether the promised gains can be delivered by mixed mode designs (Humphrey, 2013; Lynn, Uhrig, & Burton, 2010), they are nevertheless being implemented or proposed in an increasing number of longitudinal studies, such as the age 55 sweep of the UK National Child Development Survey. In such circumstances the issue of potential mode effects becomes acute.

Mode of data collection in surveys, which can range from face-to-face interviews to telephone to self-completion (paper and increasingly web), is recognised to have a potential impact on the ways in which respondents answer questions. There is now a wealth of research investigating the size and causes of these effects (de Leeuw, 2005; Dillman, 2000; Dillman & Tarnai, 1988; Holbrook et al., 2003; Jäckle et al., 2010; Kreuter, Presser, & Tourangeau, 2008; Lynn, 1998; Tourangeau & Smith, 1996). Overall, existing reviews comparing face-to-face and telephone (de Leeuw, 2005; de Leeuw & van der Zouwen, 1988) indicate that there appear to be relatively few differences between the two interviewer-administered modes of telephone and face-to-face. This contrasts with the greater mode effects found between interviewer-administered (e.g. face-to-face) and self-completion (e.g. web or paper) modes, where there are clear differences in the extent of perceived privacy of responses and interviewer control or support. Mode effects between telephone and face-to-face can, however, arise for particular types of question form and particular types of content (de Leeuw, 2005; de Leeuw & van der Zouwen, 1988; Holbrook et al., 2003; Kormendi, 1988; Tourangeau & Smith, 1996). For example, mode effects are significant for 'vaguely quantified questions' (Dillman & Tarnai, 1988) and sensitive topics such as income (Kormendi, 1988). In general, mode effects tend to be slightly larger for attitudinal or sensitive behavioural questions (Krumpal, 2013; Tourangeau, 2004; Tourangeau & Smith, 1996).

de Leeuw (2005) finds that item non response tends to be slightly higher on telephone compared to face-to-face surveys. By contrast, Biemer (2001) found that quality was higher in telephone compared to face-to-face responses, with reduced measurement error (even if higher non-response bias); and Groves (1990) argued that item non-response was lower on telephone surveys: those who respond to telephone surveys may produce slightly lower item non response because they are a more compliant group than those in face-to-face surveys.

This discrepancy in findings stems in part from the fact that apparent mode effects can capture differences in characteristics or context between those responding to the two modes, that is, factors over and above the key differences of the physical presence or absence of an interviewer. These include differences in sampling frames, response rates, survey organisation, questionnaire design and format (Jäckle et al., 2010; Roberts, 2007) as well as observed and unobserved differences between the respondents to the different modes. If such factors can be equalised across modes by a consistent survey context (the same sample and survey organisation), by random allocation to mode, and equivalent engagement of participants, then we expect (1) mode effects to be small, and (2) these mode effects would be attributable to the physical presence or absence of the interviewer only.

Direction of mode effects

Mode effects that remain after equalisation of all other conditions and context across modes will stem from the nature of the questions and the physical presence or absence of the interviewer. Mode may affect both the quality of responses and the nature of responses provided by respondents.

In relation to quality of response (that is non-response or satisficing), the visual stimulus of the interviewer's presence may have relevance for pace and engagement. Telephone interviews can increase the pace of the interview giving the impression that there is less time to respond because 'silences during telephone conversations can feel awkward' (Holbrook et al., 2003). Only hearing the question and not being able to see the interviewer during a telephone interview may also make the respondent more fatigued. This may result in more off the cuff responses (satisficing) in telephone interviews (Holbrook et al., 2003), though Jäckle et al. (2010) found no evidence of satisficing in their study of modes in the European Social Survey. In addition, a person is more likely to be multi-tasking during a telephone interview when the interviewer cannot see them. Overall questionnaire length can affect the probability of both fatigue and multi-tasking and, again, this is likely to have more of an impact on telephone interviews.

The physical presence of the interviewer may make some questions more salient and hence easier to answer, as we discuss in relation to embodied identities, below. But the physical presence or absence of the interviewer is likely to be most relevant to the nature of the response when more sensitive questions are being asked. Some questions are more at risk of respondents providing less 'open' answers because the questions (or answers) are felt to be sensitive. In such instances a face-to-face interviewer can influence the openness of responses in two ways. First, by building rapport they can encourage respondents to provide more open answers than in the more impersonal and seemingly less trustworthy context of a telephone interview. The way that face-to-face rapport enhances trust has been discussed in a number of studies (Aquilino, 1994; Cassell & Miller, 2007; DeMaio, 1984 in Roberts, 2007; Groves, 1990; Holbrook et al., 2003). Sensitive questions, however, may also be more subject to social desirability bias. That is, respondents will be inclined to give the answers that they think the interviewer will find acceptable. Social desirability bias, as a product of rapport with the interviewer, is also more likely to occur in a face-to-face interview (Krumpal, 2013). Conversational norms and the social aspects of the survey (Schaeffer & Presser, 2003) are typically stronger in face-to-face interactions (Kahn & Cannell, 1957). By contrast, the increased 'social distance' offered by telephone potentially facilitates less 'socially desirable' responses. However, without the positive influence of rapport, this may be offset by lower respondent trust in assurances of confidentiality over the phone (Groves, 1990). Therefore the direction of mode effects (face-to-face vs. telephone) in terms of which results in more open answers is not clear a priori.

However, none of these studies have tested mode effects in a longitudinal survey when rapport and trust have been established through face-to-face contact at the initial sweep. Longitudinal studies face the greatest levels of attrition between the first and second sweep. We can thus expect those who respond at the second sweep by whatever mode to have already been satisfied to some degree that their participation is worthwhile and the survey credible. In other words, from the second sweep onwards, trust should be equalised across face-to-face and telephone modes, while social desirability bias will nevertheless remain greater in face-to-face interviews.

As well as the questions themselves, the way they interact with respondent characteristics may also be important. While experimental designs should obviate the need for controls for individual characteristics, if the sample is shaped by differential response to mode then respondent characteristics may need to be controlled. The most important of these characteristics is cognitive ability. A respondent with higher cognitive ability or conscientiousness is more likely to put as much effort as is necessary to arrive at a 'correct' response (Hippler, Schwarz, & Sudman, 1987). In other words, satisficing is likely to be low regardless of interview mode. However, more cognitively demanding questions, such as attitudinal questions, may lead to respondents with lower cognitive skills satisficing or struggling to respond accurately. This indicates that care needs to be paid to ensure samples are matched on relevant characteristics before inferring mode effects.

Identity and mode effects

While there is a wealth of research on mode effects for different types of questions, there is, to date, no information on how questions to measure social and personal identity might perform in a mixed-mode

context, despite the rapid increase in research on identity in multi-topic surveys and across multiple disciplines (Casey & Dustmann, 2010; Constant & Zimmermann, 2008; Manning & Roy, 2010; Platt, 2014; Tilley, 2003). However, we can draw on general conclusions derived from the broader identity literature and what it might imply for mode effects for particular domains of identity expression to develop our theoretical expectations.

On the assumption that, other things being equal, individuals get more utility by expressing their identity (Akerlof & Kranton, 2010), we might expect responses to be largely mode invariant. However, the literature on identity measurement has emphasised responsiveness to context, that is, that people may express their identity differently in the presence of others (Hogg & Turner, 1987). On this basis, there appeared to be a reasonable case to consider that identity questions might be subject to mode effects in terms of different response patterns in the physical presence or absence of an interviewer, even under optimal implementation.

The suite of identity questions that we tested covered a wide range of different domains of identity: *profession or occupation, ethnic group, religion, national identity, political identity, family identity, father's or mother's ethnic group, marital status, gender, age & lifestage, education, and sexual identity*. Some of these, such as sexual and political identity, might be considered more sensitive while others such as life stage and gender identity could be expected to be more innocuous.

Some identity domains – ethnic group, father's/mother's ethnic group, gender, age & life stage – can all be considered embodied identities. Such identities are likely to increase in salience when they are experienced in relation to the physical social context introduced by the interviewer himself/herself. Identity and identification are well-recognised as being context contingent (Burton, Laurie, & Uhrig, 2010), social, and relational (Jenkins, 2008; Tajfel, 1981). Abrams (1996) has discussed how 'private' identities become salient and thus public in particular social situations. The interview situation represents one in which identities that may be 'private' are invited into the social or public domain through the act of questioning, and this may be reinforced by the physical presence of the interviewer.

While there has been extensive, and sometimes contradictory attention paid to the role of interviewer characteristics in determining responses (see e.g. Davis, Couper, Janz, Caldwell, & Resnicow, 2010), we might expect, in line with social identity theory (Tajfel, 1981), any face-to-face interviewer, by their physical presence, to heighten the salience of these social identities for the respondent (Krumpal, 2013); and we know that interviewer effects are greater in face-to-face compared to telephone interviews (West, Kreuter, & Jaenichen, 2013). While it might be reasonable to expect differences in ethnicity between interviewer and respondent to specifically render ethnic identity more salient, the literature on interviewer ethnicity and its impact on survey response is largely limited to the US; and even in those studies it is ambiguous and does not lead to straightforward expectations of how responses might be shaped by ethnic group (Davis et al., 2010).

Education and occupational identities are likely to increase in salience depending on the level of education and position in the occupational hierarchy. They are also potentially relational in the interaction with the interviewer, but we would expect them to be less sensitive to the physical presence of the interviewer. Similarly, family and marital status identities would appear to be little influenced by interview context.

Religious and national identities are likely to be more significant for some respondents than for others – for example for those with strong religious conviction and for those living in the smaller countries of the UK respectively. However, there would be no *prima facie* expectation that the physical presence or absence of an interviewer would heighten their salience: they are not relational in the way of physical attributes such as age and sex; and for those who have strong identities in these domains the expression of them is likely to be regarded as unproblematic in any context (Akerlof & Kranton, 2010).

Finally, there are two domains of identity that might be considered more 'sensitive' in conventional terms: these are political beliefs and sexual orientation. Voting behaviour and political orientation is notoriously susceptible to social desirability bias (Crewe, 2001; Holbrook & Krosnick, 2010; Jowell, Hedges, Lynn, Farrant, & Heath, 1993). While the identity questions are not asking the respondent

to identify their political orientation, it is plausible that any sensitivity around political beliefs will extend to the expression of identity.

Social desirability bias has also been demonstrated in relation to reporting of same-sex relationships (Villaroel et al., 2006). Clearly sexual identity covers both hetero-and homosexual identities, and it is therefore less clear whether sexual identity per se is sensitive. We might nevertheless expect *any* discussion of sexuality to be subject to the sort of conversational norms and taboos that are clearly evidenced in relation to sexual behaviours (Tourangeau & Yan, 2007), even if it is not directly linked to a specific form of sexuality. While the evidence on mode effects is mixed, if there are areas where mode effects are likely to occur, we would expect these two to be the strongest candidates.

Theoretical expectations of mode effects in measurement of identity

On the basis of this overview of relevant literature and findings, we therefore developed the following expectations:

First, we expected that 'pure' mode effects between telephone and face-to-face surveys, measured in terms of quality of responses or pattern of responses, would be limited, once we had ensured a common sample, equivalent implementation and random allocation, and controlled for differences in respondent characteristics.

Second, we expected mode differences in *quality of response*, if any, to show worse quality in telephone mode.

Third, we expected that there might be modest mode effects for particular embodied identity domains – ethnic group, parents' ethnic group, gender and age & lifestage – with more emphatic responses (stronger identification) in face-to-face compared to telephone mode, shaped by the prompt of the physical presence of the interviewer.

Fourth, we did not expect to identify mode effects across educational, occupational, religious or national identities.

And finally, fifth, we anticipated that we might find modest mode effects in relation to political and sexual identity with stronger identification among telephone respondents, since they would be less susceptible to social desirability bias.

Data, measures, design and methods

Data and sample

We exploited a unique survey methodological resource, the *Understanding Society* Innovation Panel (IP). The IP is a longitudinal survey of a representative sample of 1500 households in Great Britain. It was established in 2008 with the aim of testing methodological issues in longitudinal surveys (Jäckle & Al Baghal, 2015). Adult household members (aged 16 and over) were interviewed (for approximately 30 min) in 2008 (first wave) and every year from then on as long as they are still living within Great Britain. New members who join the households of these original sample members are also interviewed as long as they continue to live with them. Children aged 10–15 complete a self-completion youth questionnaire and are allocated the adult questionnaire when they reach 16.

We use data from the second wave of the Innovation Panel, which was in the field during March–April 2009. Responses were obtained covering a total of 1870 adults, who were interviewed either by telephone or face-to-face. The survey incorporated a range of mode experiments (Lynn et al., 2010), including that on Mixed Modes and Measurement of Identity described below. We implemented a number of sample exclusions to ensure comparability of coverage across the modes. These exclusions are shown in Table 1. First, we excluded a small number of proxy responses. To maintain random allocation, we excluded those who switched between allocated modes, discussed further below. To ensure that we had a comparable longitudinal sample, we restricted our sample to those adults who had completed an adult questionnaire at the first wave 1, or who had completed a youth questionnaire but, at age 15 were then eligible for the adult interview in the second wave.

Table 1. Sample selection.

	Number of cases excluded	Eligible sample size
Adults in wave 2 enumerated households	0	2026
Non-respondents	156	1870
Proxy respondents	82	1788
Switchers	219	1569
New entrants at wave 2 or not interviewed in wave 1	117	1452
Survey error	34	1418

Finally, we excluded 34 cases erroneously interviewed face-to-face using a telephone instrument. This resulted in a sample of 1418 cases (79% of the non-proxy respondents), 40% of whom were interviewed face-to-face ($N = 564$) and the rest by telephone ($N = 854$).

Design of the experiment and the identity module

Experimental design

The experiment on Mixed Modes and Measurement of Identity was one of a set of mode experiments conducted in wave 2 of the Innovation Panel (Lynn et al., 2010). Assignment to telephone or face-to-face mode in the experiment was random, in line with experimental practice. Households were initially randomly assigned to one of three groups. The first group was interviewed face-to-face. The second and third groups were initially assigned to telephone. However, the achieved interview mode was not completely random. For group 2, if one household member could not do the telephone interview then all household members from then on were transferred to face-to-face. For group 3, attempts were made to interview all household members by telephone, and only those who refused (soft refusals) or could not be contacted were transferred to face-to-face (McFall, 2011).

A consequence of the specific design of the experiment where we allowed non-responders among those initially assigned to a telephone mode to switch to face-to-face mode was that those interviewed face-to-face included some respondents who had earlier refused to be interviewed by telephone. Accounting effectively for differences in selection is a critical, if sometimes neglected, issue for mode evaluations (Lugtig, Lensvelt-Mulders, Frerichs, & Greven, 2011; Vannieuwenhuyze & Loosveldt, 2013). In this case, if the telephone 'refusers' (who have now become face-to-face respondents) are systematically different from telephone respondents in terms of how they respond to the identity questions then any observed mode effect will partly be a consequence of this bias. Investigating response for 'non-switchers' (Group 1 face-to-face and Group 2 and Group 3 telephone respondents) compared to 'switchers' (Groups 2 and 3 face-to-face) we found that survey response among non-switchers was higher than that of switchers by around 15 percentage points (table available on request). This shows that switchers may be more similar to non-respondents than non-switchers (see Lynn et al. (2010) for a detailed description of response rates by interview mode). For this reason it was necessary to exclude those who had switched mode (as outlined in Table 1), to stick more closely to the principles of random assignment.

Identity questions and implementation of identity experiment

We explicitly adopted a unimode design (Dillman, 2000) for the identity experiment, implementing a question structure that was intended to equalise the questionnaire context, question stimulus and cognitive burden across the modes. The identity questions themselves were modelled on a suite which had been regularly used in face-to-face mode in a cross-sectional survey, the UK government Citizenship Survey, and so had been well tested in the field. The 13 questions in the identity module related to personal identity orientation and covered the following domains: occupation/profession; ethnic or racial background; religion; national identity; political beliefs; family; father's/mother's ethnic group; marital or partnership status; gender; age and life stage; level of education; sexual orientation. For the complete questions, see Figure 1.

We'd like to know how important various things are to your sense of who you are. Please think about each thing I mention, and tell me whether you think it is important, not very important or not at all important to your sense of who you are?
READ OUT EACH AND CODE
1 Important
2 Not very important
3 Not at all important

(a) Your occupation (alt. version: profession)? INTERVIEWER: IF DK PROBE: Is that because you are retired?
(b) Your ethnic or racial background?
(c) Your religion?
(d) Your national identity?
REPEAT OF QUESTION
(e) Your political beliefs?
(f) Your family?
(g) Your father's ethnic group?
(h) Your mother's ethnic group? *{Asked only if father's ethnic group is different from mother's}*
REPEAT OF QUESTION
(i) Your marital or partnership status?
(j) Your gender?
(k) Your age and life stage?
(l) Your level of education?
(m) Your sexual orientation?

Figure 1. Identity questions in wave 2 of the *Understanding Society* Innovation Panel.

The question asked respondents the importance of each of the thirteen domains 'to their sense of who they are'. Our module differed from the Citizenship Survey questions on the exact identity dimensions measured and in the number of response categories (3 instead of 4). We reduced the number of categories in order to reduce cognitive burden and its differential impact across modes (Sudman & Bradburn, 1982, p. 269), and to avoid recency effects, where respondents are more likely to respond to the category that they have just heard. This can be an issue for telephone interviews (Holbrook, Krosnick, Moore, & Tourangeau, 2007; Krosnick, 1999). A narrow range of categories may also limit the extent of mode effects at the extremes of a scale, as found by Jäckle et al. (2010).

While a question module comprising 13 questions is somewhat long and potentially repetitive, we aimed to slow the pace and reduce satisficing by repeating the main question after every four items in the module (as illustrated in Figure 1). This was also intended to refresh the respondent's memory of what was being asked and thereby to reduce cognitive burden.

Similarly, in line with our unimode design (Dillman, 2000), we avoided using showcards so that we would not provide visual stimulus in the face-to-face interviews that was not easily available to the telephone respondents.

Within the questionnaire the identity question module appeared at the same point in both modes, thus avoiding any potential, spurious context effects. The overall questionnaire length, including the identity module, was around 30 min.

Methods

We tested mode effects by testing differences in quality of responses and pattern of responses across the two interview modes.

We measured quality in terms of item non-response by the proportion of 'don't knows' and refusals combined for each question, although there were very few refusals. We also, following Lugtig and Toepoel (2016), investigated mode differences for three other quality indicators: (i) proportions of respondents across modes who selected the same response for each of the questions in the identity module (straight-lining or satisficing), (ii) the average of the proportion of questions where the respondents selected the first category listed (primacy effect) and (iii) the average of the proportion of questions where the respondent selected the last category listed (recency effect).

We tested whether there were any significant differences in the proportions of responses that might indicate response quality differences across the modes in line with our expectations, using standard statistical tests: t-tests for continuous variables and, for discrete variables, Pearson's χ^2 test as well as Fisher's exact test.[1]

We tested differences in patterns of responses across modes by estimating a series of ordered logit models including mode as an explanatory variable for each of the identity domains, which allowed us to take account of the ordinal nature of the response categories.

In the implementation of the mode experiment, our aim was to ensure comparability of respondent characteristics across modes through random assignment (Heerwegh, 2009). However, differences in inclusion and response across modes meant there was still potential for the two groups to differ on relevant characteristics, through the exclusion of switchers (discussed above) and differences in rates of non-response or different types of people being non responders across the two modes; and randomly occurring differences, though these latter should be adequately addressed by the experimental design.

We therefore drew on key socio-demographic measures in order to compare – and control for the characteristics of the two samples. As discussed earlier it is particularly important to control for cognitive ability, as identity questions, being both attitudinal and autobiographical are more cognitively demanding than 'factual' questions (Hippler et al., 1987; Tourangeau, Rips, & Rasinski, 2000). As there was no direct measure of cognitive ability we included educational qualifications, which provide a rough approximation.

We describe the control variables used in the analyses in the Appendix 1. Table 2 provides the values and distributions of these socio-demographic variables across the two mode samples. Though the differences are not large, we can nevertheless observe some systematic variation.

We therefore estimated both an unadjusted model and an adjusted model controlling for the observed characteristics described in Table 2. For parsimony and ease of interpretation, we simply report the predicted probabilities for the responses by mode, but full tables of results are available on request.

While interviewer characteristics, such as ethnic group, may be relevant to the experienced salience of the identity domains, we lacked complete information on interviewer characteristics and so could not control for these. At the same time, ethnic minorities and immigrants are only present in very small numbers in our study, so mode effects cannot be estimated separately for them.

In all analyses we estimate robust standard errors to account for clustering of participants within areas. As the IP sample was selected with an almost equal selection probability (design weight for 99.31% of the sample is 1), we do not use design weights in the analysis.

Results

Quality of response

We did not find any statistically significant differences across modes for straightlining, primacy or recency effects (results available on request). For item non-response, we found that interview mode was only associated with occupation. Counter to our expectations, the probability of item non-response was higher in face-to-face mode (see Table 3). Looking further into the characteristics of these non-responders we found, however, that those in our analysis sample who reported 'don't know' to the occupational identity question were not employed. The greater tendency of non-employed to say 'don't

Table 2. Comparison of socio-demographic characteristics of face-to-face and telephone samples.[a]

	Face-to-face	Telephone	χ^2 test p-value
Female?	56.7%	55.7%	0.71
Born in UK?	92.7%	94.0%	0.33
Ethnic group is non-White?[b]	5.9%	4.9%	0.44
Has access to a car?	81.4%	85.9%	0.02
Number of own children in household	0.4	0.4	
Birth cohort			
1900–1940	21.3%	20.1%	0.61
1941–1950	16.5%	17.1%	0.77
1951–1960	15.4%	20.4%	0.02
1961–1970	20.6%	20.1%	0.85
1971–1980	15.6%	12.5%	0.10
1981–1990	7.1%	7.3%	0.91
1991–2000	3.5%	2.5%	0.23
Educational qualifications			
First degree or higher	21.6%	24.1%	0.27
School completion	32.7%	32.4%	0.90
Other	26.4%	27.3%	0.72
No qualifications	19.3%	16.2%	0.14
Owns house?	77.8%	81.6%	0.08
Religion			
None	47.0%	47.0%	0.98
Christian	45.4%	46.6%	0.66
Other religion	7.5%	6.4%	0.41
Current health status			
Very good health	50.3%	52.3%	0.45
Good Health	46.5%	40.7%	0.03
Poor Health	3.2%	7.0%	0.00
Marital Status			
Never Married	15.1%	13.5%	0.41
Married	65.2%	70.9%	0.02
Separated, divorced or widowed	19.7%	15.5%	0.04
Current activity status			
Employed	55.9%	59.3%	0.21
Unemployed	3.2%	2.5%	0.41
Retired	27.9%	27.3%	0.80
Other	12.9%	11.0%	0.26
Region of residence			
London	9.4%	7.4%	0.18
East and South	31.7%	36.5%	0.06
Midlands	16.5%	16.8%	0.88
North	30.3%	24.8%	0.02
Wales	5.7%	5.9%	0.87
Scotland	6.4%	8.6%	0.13
Number of observations	564	854	

[a]Excludes respondents who switched interview modes from their initial assignments.
[b]Aggregate of all 2011 ethnic group categories, where respondents do choose one of the White groups.

know' was consistent with overall lower reported importance of occupation among non-employed who did provide a substantive response to this question. For those not in paid work, occupational identity is less relevant and so less likely to comprise a pre-formed concept for the respondent. The proportions of non-employed were higher among the face-to-face sample than the telephone sample (6.5% as compared to 2.8%) and these differences were statistically significant suggesting that the slightly greater proportion of non-response in the face-to-face sample was being driven by the higher proportion of non-employed. The overall findings imply that item non-response is not a major concern for identity questions.

Response patterns

Tables 4 and 5 illustrate the probabilities of discrete responses by mode across the identity domains from the estimates of our ordered logit models with and without controls. In general, mode appears to

Table 3. Tests of differences in item non-response by interview mode.[a]

	Item non-response[b]			Tests of mode effects		Number of item non-response cases in combined sample (N = 1418)
				χ^2 test	Exact Fisher's test	
	F2F (%)	Telephone (%)	All (%)	p-values	p-values	
Occupation/profession	*3.02*	*1.17*	*1.91*	*0.013*	*0.016*	*27*
Ethnic or racial background	3.48	1.46	2.29	0.821	1.000	11
Religion	2.55	0.90	1.53	0.673	0.651	4
National identity	0.71	0.82	0.78	0.696	1.000	9
Political beliefs	0.36	0.23	0.28	0.176	0.222	6
Family	0.53	0.70	0.64			0
Father's ethnic group[c]	0.71	0.23	0.42	0.061	0.073	16
Marital or partnership status	0.00	0.00	0.00	0.097	0.168	9
Gender	1.78	0.70	1.13	0.101	0.125	22
Age and life stage	1.07	0.35	0.64	0.770	0.747	9
Education	0.89	1.99	1.55	0.766	1.000	2
Sexual orientation	0.71	0.59	0.64	0.372	0.492	20

Note: Statistically significant differences at the 5% level are indicated in italics.

[a]Excludes respondents who switched interview modes from their initial assignments.

[b]Item non-response is measured by the proportion of respondents who respond to these questions with 'don't know' or 'refuse'. Most of these are 'don't know' responses, with only a handful of 'refusals'.

[c]'Your mother's ethnic group' was asked of only those whose mother's race/ethnicity was different from that of her father and so there were only a handful of cases who were asked this question. So, we have omitted this question from the analysis.

Table 4. Estimates of identity response from ordered logit using only mode effect as covariate.[a]

			Predicted probability of identity responses					
	Estimated coefficient of inter-view mode	p-value	Face-to-face			Telephone		
			Important	Not very important	Not at all important	Important	Not very important	Not at all important
Occupation/profession	0.09	0.46	0.66	0.17	0.16	0.68	0.16	0.15
Ethnic or racial background	−0.15	0.18	0.56	0.30	0.14	0.52	0.32	0.16
Religion	−0.09	0.47	0.38	0.36	0.26	0.36	0.37	0.27
National identity	−0.15	0.28	0.75	0.21	0.05	0.72	0.23	0.06
Political beliefs	**0.38**	**0.00**	**0.34**	**0.46**	**0.21**	**0.43**	**0.42**	**0.15**
Family	−0.38	0.34	0.98	0.01	0.01	0.97	0.02	0.01
Father's ethnic group[b]	0.06	0.61	0.47	0.32	0.21	0.48	0.32	0.20
Marital or partnership status	0.01	0.95	0.72	0.16	0.12	0.72	0.16	0.12
Gender	−0.24	0.10	0.80	0.14	0.06	0.76	0.16	0.08
Age and life stage	−0.19	0.11	0.62	0.31	0.08	0.57	0.34	0.09
Education	0.22	0.13	0.71	0.24	0.05	0.75	0.20	0.04
Sexual orientation	−0.09	0.46	0.65	0.24	0.10	0.63	0.25	0.11
Occupation	0.15	0.37	0.68	0.16	0.16	0.71	0.15	0.14
Profession	0.04	0.80	0.65	0.18	0.17	0.66	0.18	0.16

Note: Mode effects that are statistically significant at 1% level of significant are indicated in bold.

[a]Excludes respondents who switched interview modes from their initial assignments. Predicted probabilities have been computed by averaging the predicted probabilities of all observations.

[b]'Your mother's ethnic group' was asked of only those whose mother's race/ethnicity was different from that of her father and so there were only a handful of cases who were asked this question. So, we have omitted this question from the analysis.

have little association with responses. In the unadjusted analysis, mode differences were found to be significant at the 1% level for political beliefs and at around the 10% level for gender, age & life stage and education. For political beliefs, respondents to the telephone mode were more likely to consider

Table 5. Estimates of identity response from ordered logit using with mode effect and additional characteristics as covariates.[a]

| | Estimated coefficient of interview mode | p-value | Predicted probability of identity responses | | | | | |
| | | | Face-to-face | | | Telephone | | |
			Important	Not very important	Not at all important	Important	Not very important	Not at all important
Occupation/ profession	0.09	0.53	0.69	0.18	0.13	0.71	0.17	0.12
Ethnic or racial background	−0.15	0.22	0.57	0.30	0.13	0.53	0.32	0.15
Religion	−0.13	0.36	0.33	0.50	0.17	0.30	0.51	0.19
National identity	−0.12	0.38	0.76	0.20	0.04	0.74	0.22	0.05
Political beliefs	**0.34**	**0.00**	**0.34**	**0.47**	**0.19**	**0.42**	**0.44**	**0.14**
Family	−0.38	0.43	1.00	0.00	0.00	1.00	0.00	0.00
Father's ethnic group b	0.08	0.48	0.47	0.34	0.19	0.49	0.33	0.17
Marital or part- nership status	−0.08	0.61	0.80	0.14	0.05	0.79	0.15	0.06
Gender	−0.22	0.14	0.81	0.13	0.06	0.78	0.16	0.07
Age and life stage	−0.17	0.16	0.62	0.31	0.07	0.58	0.34	0.08
Education	0.11	0.45	0.77	0.20	0.03	0.79	0.18	0.03
Sexual orienta- tion	−0.13	0.30	0.67	0.24	0.09	0.64	0.26	0.10
Occupation	0.23	0.27	0.69	0.17	0.13	0.74	0.15	0.11
Profession	0.00	0.99	0.68	0.19	0.12	0.68	0.19	0.12

Note: Mode effects that are statistically significant at 1% level of significance are indicated in bold.

[a]Excludes respondents who switched interview modes from their initial assignments. Additional covariates included in the model are those listed in Table 2. Predicted probabilities have been computed by averaging the predicted probabilities of all observations.

[b]'Your mother's ethnic group' was asked of only those whose mother's race/ethnicity was different from that of her father and so there were only a handful of cases who were asked this question. So, we have omitted this question from the analysis.

their beliefs important, and this was also the direction of the effect for education, while for gender and age & lifestage it was face-to-face respondents who were more likely to judge them as important. While the results for gender and age & lifestage and political identity were in line with our expectations, those for education were not.

When re-estimating the ordered logit controlling for characteristics (Table 5), the coefficient for mode effects for education reduced by half and was rendered statistically insignificant. The apparent mode effect could thus be convincingly attributed to differences in characteristics between the two samples. While the effects for gender and age & life stage also reduced in statistical significance, the coefficients reduced much less, suggesting that, in a larger sample, there might be evidence of a slight mode effect, and one that was in the expected direction. The effect was, however, small in magnitude: predicted probabilities indicated around a three percentage point increase in the chances of finding these aspects of identity 'important' in the face-to-face sample. Overall, it appeared that design and adjustment effectively compensated for any spurious mode effects and that there was little evidence for mode differences across 12 of the 13 domains.

The mode effect for political beliefs remained very similar across the two models in both size and statistical significance. The probability of considering political beliefs as 'important' only reduced to being eight rather than nine percentage points higher in telephone rather than face-to-face mode, once demographics were controlled, with over 40% of telephone responders but only about a third of face-to-face responders evaluating their political beliefs as important. The difference is sufficient to suggest sensitivity to mode for political identity, in line with the existing findings on greater openness about political *beliefs* in telephone surveys. However, with a battery of 13 questions, we cannot dismiss the possibility that one such significant association could be due to chance.

Discussion and conclusions

In this paper, we tested for mode effects for a varied module of 13 identity questions in a nationally representative longitudinal survey sample. We utilised a specifically longitudinal sample for two reasons: first, to recognise the increasing dominance of large longitudinal surveys in social research, and second the relative paucity of analysis of mode effects in such studies. By equalising survey context, questionnaire design and respondent characteristics, through random assignment and through operationalising a unimode design, we aimed to establish whether any of the identity questions were subject to mode effects operating purely through the differences engendered by the physical versus aural only presence of an interviewer. Because we were working with an established sample (that is those who chose to respond to the survey a second time), we were able to assume a certain level of trust in and engagement with the study from respondents, regardless of current mode of interview.

Of our 13 identity domains, none showed variation in quality as measured by non-response or don't know, straightlining, primacy and recency effects and only one domain, political identity, showed clear evidence of mode differentials on substantive response. In telephone mode, respondents were more likely to perceive their political beliefs as 'important' to their identity, with a difference of around eight percentage points. This is consistent with the literature that expression of political beliefs and behaviour are susceptible to social desirability bias, and, in a situation of equal engagement with and trust in the survey, that telephone may offer greater scope for expressing political views (Cassell & Miller, 2007; Holbrook & Krosnick, 2010). This may be due to the reduced influence of those social norms that are created through rapport (Cappella, 1990; Cassell & Miller, 2007). While we were not asking about political beliefs, sensitivity to the general domain of 'politics' is arguably also likely to apply to political identity. However, we did not find a comparable effect for the sexual orientation question that we also considered 'sensitive'. The fact that we found little evidence for mode effects, including for those domains where we did not anticipate strong differences across modes provides some endorsement for our unimode approach, suggesting that we had successfully eliminated spurious sources of variation across modes. This may be of practical value for those intending to implement such questions in mixed mode surveys.

We conclude that identity questions do not show strong evidence of direct mode effects. In a setting designed to eliminate or minimise sources of spurious variation across modes, the allocation of respondents to either telephone or face-to-face interview should not be expected to impact their responses to identity questions, and hence allows for comparison both within and across survey sweeps. While our findings have focused on the longitudinal context where mode may change across as well as within sweeps, and where we can condition on initial engagement, our findings may also have implications for cross-sectional surveys utilising different modes, if measures of trust or engagement within the survey can be exploited to adjust for potential variation. While our findings indicated, in line with studies of political beliefs, that political identity may be more sensitive to expression in the physical presence of the interviewer, we would treat this finding with caution until it can be supported by further studies.

Note

1. While Fisher's exact test is more accurate when the size of at least one cell in the two-way frequency table of the two dichotomous variables is less than five, it is also a more conservative test. See Little (1989) for discussion of this debate. So, here we report both tests.

Acknowledgements

Understanding Society: Innovation Panel, Waves 1-5, 2008-2012 [computer file]. 2nd Edition. Colchester, Essex: UK Data Archive [distributor], September 2012. SN: 6849.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Alita Nandi is a research fellow at ISER, University of Essex. She conducts empirical research primarily in areas of ethnicity, identity and gender specifically looking at well-being (subjective as well as economic), labour market activities and partnership formation and dissolution. She has published in Ethnic and Racial Studies, Applied Economics, Longitudinal and Life Course Studies. She co-authored "A Practical Guide to Using Panel Data" with S. Longhi in 2014 (SAGE publications Ltd).

Lucinda Platt is Professor of Social Policy and Sociology at the London School of Economics and Political Science. Her research interests focus on immigration, ethnicity, poverty, disability and gender. She has published widely in these areas in journals including Ethnic and Racial Studies, Social Forces, Sociology and Oxford Economic Papers. In 2011, she published Understanding Inequalities: Stratification and Difference (Polity); and 2016 saw the publication of her co-edited volume (with Hartley Dean), Social Advantage and Disadvantage (OUP).

References

Abrams, D. (1996). Social identity, self as structure and self as process. In W. P. Robinson (Ed.), *Social categories and identities: Developing the legacy of Henri Tajfel* (pp. 143–168). Oxford: Butterworth-Heinemann.

Akerlof, G., & Kranton, R. (2010). *Identity economics: How our identities shape our work, wages and well-being.* Princeton, NJ: Princeton University Press.

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

Biemer, P. P. (2001). Non response bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics, 17*, 295–320.

Burton, J., Laurie, H., & Uhrig, S. C. N. (2010). *Understanding Society Innovation Panel Wave 2: Results from methodological experiments* (Understanding Society Working Paper 2010-04). Colchester: University of Essex.

Cappella, J. N. (1990). On defining conversational coordination and rapport. *Psychological Inquiry, 1*, 303–305.

Casey, T., & Dustmann, C. (2010). Immigrants' identity, economic outcomes and the transmission of identity across generations. *The Economic Journal, 120*, F31–F51.

Cassell, J., & Miller, P. (2007). Is it self-administration if the computer gives you encouraging looks? In F. G. Conrad & M. F. Schober (Eds.), *Envisioning the survey interview of the future* (pp. 161–178). New York, NY: Wiley.

Constant, A., & Zimmermann, K. (2008). Measuring ethnic identity and its impact on economic behavior. *Journal of the European Economic Association, 6*, 424–433.

Couper, M. P. (2011). The future of modes of data collection. *Public Opinion Quarterly, 75*, 889–908.

Crewe, I. (2001). The opinion polls: Still biased to labour. *Parliamentary Affairs, 54*, 650–665.

Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research, 25*, 14–26.

de Leeuw, E. (2005). To mix or not to mix data collection modes in surveys. *Journal of Official Statistics, 21*, 235–255.

de Leeuw, E., & van der Zouwen, J. (1988). Data quality in telephone and face-to-face surveys: A comparative meta-analysis. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 283–300). New York, NY: Wiley.

DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying subjective phenomena* (Vol. 2, pp. 257–282). New York, NY: Russell Sage Foundation.

Dillman, D., & Tarnai, J. (1988). Administrative issues in mixed mode surveys. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 509–528). New York, NY: Wiley.

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

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the internet. *Social Science Research, 38*, 1–18.

Groves, R. M. (1990). Theories and methods of telephone surveys. *Annual Review of Sociology, 16*, 221–240.

Heerwegh, D. (2009). Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research, 21*, 111–121.

Hippler, H.-J., Schwarz, N., & Sudman, S. (1987). *Social information processing and survey methodology*. New York, NY: Springer-Verlag.

Hogg, M. A., & Turner, J. C. (1987). Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology, 26*, 325–340.

Hogg, M. (2006). Social identity theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 111–136). Palo Alto, CA: Stanford University Press.

Holbrook, A. L., Green, M. C., & Krosnick, J. A. (2003). Telephone versus face-to-face interviewing of national probability samples with long questionnaires. *Public Opinion Quarterly, 67*, 79–125.

Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly, 74*, 37–67.

Holbrook, A. L., Krosnick, J. A., Moore, D., & Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly, 71*, 325–348.

Humphrey, A. (2013, July). *European Social Survey mixed mode experiment*. Presentation to the European Survey Research Association conference, Ljubljana.

Jäckle, A., & Al Baghal, T., with Burton, J., & Lynn, P. 2015. *Understanding Society, The UK Household Longitudinal Study Innovation Panel, Waves 1–7, User Manual. Version 1.0*. Colchester: Institute for Social and Economic Research, University of Essex.

Jäckle, A., Roberts, C., & Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review, 78*, 3–20.

Jenkins, R. (2008). *Social identity* (3rd ed.). London: Routledge.

Jowell, R., Hedges, B., Lynn, P., Farrant, G., & Heath, A. (1993). Review: The 1992 British Election: The failure of the polls. *Public Opinion Quarterly, 57*, 238–263.

Kahn, R. L., & Cannell, C. F. (1957). *The dynamics of interviewing: Theory, technique and cases*. New York, NY: Wiley.

Kormendi, E. (1988). The quality of income information in telephone and face to face surveys. In R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, II, & J. Waksberg (Eds.), *Telephone survey methodology* (pp. 341–356). New York, NY: Wiley.

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly, 72*, 847–865.

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

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality and Quantity, 47*, 2025–2047.

Little, R. (1989). Testing the equality of two independent binomial proportions. *The American Statistician, 43*, 283–288.

Lugtig, P., Lensvelt-Mulders, G. J., Frerichs, R., & Greven, A. (2011). Estimating nonresponse bias and mode effects in a mixed-mode survey. *International Journal of Market Research, 53*, 669–686.

Lugtig, P., & Toepoel, V. (2016). The use of PCs, smartphones, and tablets in a probability-based panel survey: Effects on survey measurement error. *Social Science Computer Review, 34*(1), 78–94.

Lynn, P. (1998). Data collection mode effects on responses to attitudinal questions. *Journal of Official Statistics, 14*(1), 1–14.

Lynn, P., Uhrig, S. C. N., & Burton, J. (2010). *Lessons from a randomised experiment with mixed-mode designs for a household panel survey* (Understanding Society Working Paper: 2010-03). Colchester: University of Essex.

Manning, A., & Roy, S. (2010). Culture clash or culture club? National identity in Britain. *The Economic Journal, 120*, F72–F100.

McFall, S. L. (Ed.). (2011). *Understanding Society – The UK Household Longitudinal Study, Innovation Panel, Waves 1 and 2, User Manual*. Colchester: University of Essex.

Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin, 108*, 499–514.

Platt, L. (2014). Minority groups' national, ethnic and religious identity: Investigating generational change. *Ethnic and Racial Studies, 37*, 46–70.

Roberts, C. (2007). *Mixing modes of data collection in surveys: A methodological review* (ESRC National Centre for Research Methods Briefing Paper, NCRM Methods Review Papers NCRM/008). Southampton: University of Southampton.

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65–88.

Sudman, S., & Bradburn, N. M. (1982). *Asking questions*. San Francisco, CA: Jossey-Bass.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge: Cambridge University Press.

Tilley, J. (2003). Party identification in Britain: Does length of Time in the electorate affect strength of partisanship? *British Journal of Political Science, 33*, 332–344.

Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775–801.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.

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

Tourangeau, R. & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883.

Vannieuwenhuyze, J. T. A., & Loosveldt, G. (2013). Evaluating relative mode effects in mixed-mode surveys: Three methods to disentangle selection and measurement effects. *Sociological Methods and Research, 42*, 82–104.

Verkuyten, M. (2007). Religious group identification and inter-religious relations: A study among Turkish-Dutch Muslims. *Group Processes & Intergroup Relations, 10*, 341–357.

Villaroel, M., Turner, C., Eggleston, E., Al-Tayyib, A., Rogers, S., Roman, A., Cooley, P., & Gordek, H. (2006). Same-gender sex in the United States: Impact of T-Acasi on prevalence estimates. *Public Opinion Quarterly, 70*, 166–196.

West, B. T., Kreuter, F., & Jaenichen, U. (2013). 'Interviewer' effects in face-to-face surveys: A function of sampling, measurement error, or nonresponse? *Journal of Official Statistics, 29*, 277–297.

Appendix 1

The control variables used in the analysis are described below. All variables were measured in wave 2, except for time-invariant variables which were measured the first time the respondent was interviewed.

Birth cohort: Based on their years of birth, respondents were categorised into 8 birth cohorts: 1900–1940, 1941–1950, 1951–1960, 1961–1970, 1971–1980, 1981–1990, 1991–2000, 2001–2010. The reference category was 1900–1941.

Gender: This variable was 1 for female, 0 for male

Migrant status: Respondents were asked whether they were born in any of the 4 UK countries or outside UK. This variable was 1 for those who reported they were born in any one of 4 UK countries, 0 otherwise.

Ethnic group: Respondents were asked what their ethnic group was and were asked to choose from 15 response (or 14 if in Scotland) options. This question was a variant of the 2001UK Census ethnic group question. All those who chose 'white British', 'White Scottish', 'Other White British', were coded as 1, all others 0.

Religious affiliation: Respondents were asked whether they belonged to a religion. If they said yes, they were asked about their religious affiliation (11 response options). Those who said they did not belong to a religion they were asked if they were brought up in one and if so what that was (11 response options). These two sets of responses were combined and a combined religious affiliation variable was constructed. Their religious affiliation was recoded into a 3 category variable: No religious affiliation, Christian, Other religion. The reference category was Christian.

Region of residence: The region of residence of the respondent was coded into 6 category variable: London, East and South excluding London, Midlands, North, Wales and Scotland. London was the reference category.

Highest educational qualification: This was categorised into these categories: First degree or higher, School completion, Other, No qualifications. The reference category was no qualifications.

Main activity status: Respondents were asked about their current activity status and they could choose from 11 categories. This was recoded into a 4-category variable: employed (in paid employment or self-employed), unemployed, retired, other (full-time student, taking care of family, on maternity leave, long term ill or disabled, in government training). The reference category was employed.

Marital status: The current marital status was categorised into a 3 category variable: Never married, Married, in civil partnership or cohabiting and if separated, divorced or widowed. The reference category was Married, in civil partnership or cohabiting.

Health: Respondents were asked about their general health and they could choose one of 5 categories which were recoded into a 3 category variable: Excellent or very good, Good or fair and poor. The reference category was Excellent or very good.

House ownership: This is a 0–1 variables, which equals 1 if the house that the household resides in is owned (mortgaged or owned outright or part-owned & part-rented) by the household and 0 otherwise (that is, if rented privately or socially or some other arrangement).

Car access: In the household questionnaire, the household respondent was asked whether they or anyone in the household had access to a car or van for private use. Additionally, the adult individual respondents who reported they had a driving licence were asked if they normally had access to a car or van which they could use whenever they wanted to. These variables were combined into a two category variable – 1 if the individual or the household have access to a car, 0 otherwise.