

How Face-to-Face Interviews and Cognitive Skill affect Non-response: A randomized experiment assigning mode of interview

Andrew Gooch,
Department of Political Science, University of California Los Angeles
aagooch@ucla.edu

and

Lynn Vavreck,
Department of Political Science, University of California Los Angeles
lvavreck@ucla.edu*

September 8, 2014

*This research is supported by a grant from the National Science Foundation (SES-1023940). We thank R. Brian Law, who kept things running on time and effectively at the MGM Grand, and all of the graduate students who helped administer the experiment in Las Vegas: Michael LaCour, Felipe Nunes, Sylvia Friedel, Gilda Rodriguez, Adria Tinnin and Chris Tausanovitch. Doug Rivers and Jeff Lewis provided programming support; and, John Aldrich, Larry Bartels, Alan Gerber, Gary Jacobson, Simon Jackman, Vince Hutchings, Gary Segura, John Zaller, and Brian Humes helped with the design of the experiment at a meeting sponsored by the National Science Foundation in the summer of 2010.

Abstract

For decades, projects like the American National Election Study have used in-person interviews to collect data. Increasing costs have led investigators to consider abandoning the face-to-face interview, but little is known about the effects of the transition. We offer evidence on item non-response from an experiment in which an adult population was randomly assigned to an in-person or self-complete survey after agreeing to participate. For nearly every topic, we find more item non-response in the face-to-face mode. This difference is exacerbated for respondents with low levels of cognitive abilities and best illustrated by differences to fact-based questions, which respondents are more likely to avoid answering in the in-person setting and more likely to get correct in the self-complete interview. Moving from high to low levels of cognitive ability, an otherwise average respondent is five times more likely to say “don’t know” in a face-to-face interview than in a self-completed survey.

For more than 60 years, most of the data used in political science research on elections, public opinion, and parties was gathered by the American National Election Study (ANES) through in-person, face-to-face surveys. Respondents for ANES surveys are typically drawn from household samples selected via a multi-stage, clustered area probability sampling methodology. Near complete coverage of the population by the sampling frame, high response rates (upwards of 80% in the 1950s-1960s), and the willingness of respondents to submit to hours of questioning (characterized as a function of the “rapport” built between respondent and interviewer) are the main reasons this approach is known as the “gold standard” of survey research (Scioli and Granato 2003).

In-person interviewing, however, is expensive — and becoming more so. In 2012, it cost approximately \$2,100.00 per case (inclusive of both interviews but exclusive of staffing costs) to produce the ANES (Hutchings et al 2010). Compounding the costs of the interviews is the design effect of the project, which reduces the effective sample size by roughly 30% due to the clustering of cases (and the potential unobserved correlation among respondents in the same clusters).¹ The reduction in effective sample size increases the actual cost per case to just under \$2,900.00. Staffing costs may add an additional \$1500.00 to the cost per completed case.

The ANES time series is the nation’s best data on political attitudes and behavior. It has collected data on a set of core variables every four years since 1948, and until 2006 the frequency was every two years. As costs continue to rise and budgets shrink, projects like the ANES have had to make tough choices: decrease the number of interviews (already artificially diminished by the design effects) or abandon the gold-standard attributes that makes the project one of the highest quality, federally funded social science surveys in the field. An attractive alternative to face-to-face interviewing is online, self-complete surveys because of the proliferation of Internet usage, the increase in computer literacy, and the

¹The ANES project staff estimate the design effect for the 2012 in-person study to be, on average, 1.9 (ANES 2013, p. 33).

lower cost per interview. The ANES has recently included on-line, self-completed data in its deliverable, but has not yet abandoned the face-to-face interview.

While on-line, self-complete surveys are growing in popularity, much of the research fails to address the mode differences between face-to-face interviewing and self-complete surveying – and where survey mode has been addressed, it is rarely through randomized experimentation after the acceptance of an invitation to complete a survey. Regardless of which survey mode is preferred by researchers, budgeting realities are now forcing social scientists to pursue self-completed surveys as an option. As a result, it is important to understand the possible response differences brought about by mode alone (holding sampling constant), especially if comparisons are to be made to previous surveys using other modes of interview. In this paper, we offer findings on one of the differences: the rate at which respondents say that they “do not know” the answers to survey questions or simply skip questions – and how a person’s level of cognitive ability interacts with the survey mode to produce different results across the two modes of inquiry.

To isolate the effects of mode, we randomly assigned 1,010 adult respondents to complete either a face-to-face or self-complete survey with identical questions ($N = 505$ per mode). These treatment assignments occurred *after* respondents agreed to participate in the survey. This feature of the experimental design is critical as it allows for a test purely of differences due to mode of interview, not one that conflates sampling method, response rates, or sampling frame with mode effects. This unique design eliminates any confounds related to sampling or selection bias associated with the mode of survey administration. The design isolates differences due to mode and helps illustrate the ways in which answering questions with an interviewer in the room generates different responses, on average, than answering questions on a computer, nearly all else equal.

Survey Mode and Non-response

The theoretical underpinnings of survey non-response can be characterized generally by cognitive elaboration models (Schwarz et al 1992; 1994; Sudman et al 1996) and more specifically by satisficing theory (Krosnick 1991), both of which broadly assume that a survey response requires a sufficient level of cognitive skill or motivation. Respondents first decipher the meaning and purpose of a question, find relevant information in their minds, condense that information into a summary report, and then try to make that summary report fit within the limitations of the question design (see also Schwarz and Bonner 2001; Tourangeau, Rips, Rasinski 2000; Krosnick 1991; Stack, Schwarz and Wanke 1991). For many respondents, this response process comes naturally, but for subsets of the sample, a shortcut might be the easiest response to a question deemed difficult, which can lead to satisficing, primacy, or recency effects (Oppenheim 1992; Krosnick 1991; Feick 1989; Bradburn and Sudman 1988; Converse 1976; Coombs and Coombs 1976). Influences on these effects include the respondent's ability and motivation, and the difficulty of the question (Carmines and Stimson 1980; Krosnick 1991; Mondak and Davis 2001), but satisficing, specifically with a non-response, is also attributed to lower cognitive skill (Krosnick 1992; Krosnick, Holbrook, Bernet, Carson, Hanemann, Kopp, Mitchell, Press, Ruud, Smith, Moody, Green and Conaway 2002). Socio-economic characteristics may also contribute to non-response including race, gender, income, and education (Althaus 2003; Converse 1976; Francis and Busch 1975; Rapoport 1982).

Given the underpinnings of non-response, there are many reasons to expect both response and non-response differences due to survey mode, and a small group of scholars have investigated these differences across various modes of interview with fruitful results (Chang and Krosnick 2010; Heerwegh and Loosveldt 2008; Heerwegh 2009; Acree, Ekstrand, Coates and Stall 1999; Fowler, Roman and Xiao 1998; De Leeuw 1992; Gano-Phillips and Fincham 1992; Bishop, Hippler, Schwarz and Fritz 1988; Kiesler and Sproull 1986; Sudman and Bradburn 1974). No study to date, however, has directly compared

in-person interviewing to self-completed surveys using an adult population with random assignment to mode *after* respondents agree to participate. As a result, we know very little about non-response differences across these two modes in non-student populations. Chang and Krosnick (2010) offer the only other post-acceptance randomized mode experiment comparing these two modes. They randomly assigning college students to complete a survey on a computer or answer questions asked by an interviewer heard over an intercom. Their findings, albeit on a student population, motivated our search for interactions based on cognitive skill levels, as they find students with lower standardized test scores (information gleaned from their admissions files) had a more difficult time answering questions in the interviewer-assisted intercom mode. A few other studies compare outcomes of in-person interviews to outcomes from online surveys, but despite their notoriety, these projects are not randomized experiments and thus cannot separate the effects of survey mode from sampling methodology (Malhotra, Krosnick 2007; Sanders, Clarke, Stewart, and Whiteley 2007).

A Randomized Mode Experiment

In order to isolate response differences due only to mode, we randomly assigned respondents to complete surveys via one of two different methods *after* they agreed to participate in a survey. Waiting until they have agreed to participate ensures that any observed differences between modes are not due to differences in the types of people who agree to participate via different modes. We conducted the experiment at Television City, the CBS research facility within the MGM Grand Hotel in Las Vegas, Nevada. CBS uses this facility to conduct daily focus groups on its programming. According to CBS, over a million people from all 50 states pass by the research facility in a given year. Because of the ready access to an adult population and CBS's demonstrated record of recruiting participants to take surveys, Television City was an appealing choice for this research. More importantly, however, CBS built several staged living rooms in the research facil-

ity in order to test people’s reactions to 3-D TVs. These living rooms made a perfect setting for the in-person interview as they mimic the “at home” environment in which ANES interviews are completed. Sitting in a living room, even if it is not your own living room, across the coffee table from an interview, provides an interview experience that more closely resembles an ANES interview than sitting in a small conference room at a desk. Of course this setting is not perfectly reflective of all the ways the ANES in-person interviews are done, but staging a living-room and conducting the face-to-face interviews in it struck us as a step in the right direction in terms of replicating the ANES experience, especially relative to a sterile conference or classroom setting. Self-completed surveys were done on desktop computers in private rooms adjacent to the living room spaces.

In exchange for participation in the study, respondents received a \$5 gift card to one of the many Starbucks Coffee cafes within the MGM Grand. The face-to-face interviews were conducted by six professional interviewers who were trained by two members of the ANES Board of Advisorys in advance of the project. The experiment ran from July 26th to August 6th, 2011. The research facility was open from 10 AM to 7 PM every day and we processed roughly 75-100 respondents each day. Respondents were recruited from three areas within the hotel by CBS and project staff using the CBS protocols for their focus group projects. The experiment was described as a survey on current events for a university-sponsored research project called the State of the Nation Project — the mode of interview was not mentioned. The MGM Grand placed large banners advertising the project at these locations to attract potential respondents and staff members distributed cards to people as they walked through the three areas inviting people to “stop in, charge their phone, get a Starbucks gift card, and participate in a short survey.”

We randomized people into conditions by blocking on three demographic indicators: age, race, and sex. The blocking design ensured that demographic covariates were not collinear with the mode assignments (see Green and Gerber 2012, p. 72-79 for a complete description of this method). By randomizing within each block, we eliminated the

possibility that, just by chance, an imbalance of demographics occurred across treatment groups.² Respondents were grouped according to observed demographics by research staff after agreeing to participate in the survey but before randomization. We sorted respondents into three age groups (below 30, 31 to 59, and 60 plus), three race categories (white/Asian, African American, and Hispanic), and sex.³

An Overview: Differences by Mode and Question Type

We start by providing an overview of non-response for every question in the survey. Figure 1 shows the percentage-point difference with 95% confidence intervals by mode for all questions. To estimate the treatment effect we subtract the self-complete non-response from the in-person non response, thus positive values indicate higher levels of non-response in the face-to-face treatment. The survey was written to cover many different types of questions, topics, and styles of presentation (grids, open-ends, multiple choice, etc . . .) Question topics represent four broad areas: factual knowledge, behavioral reports, issue positions, and attitudes/evaluations. The questions cover both political and non-political topics and are generally classified as easy or difficult.⁴ The survey was also written to

²As a means of reducing sampling variability, a blocked design also allows for calculating standard errors within each block for difference of means estimations. All standard errors reported here are calculated taking account of the blocking in this manner. For a detailed explanation of blocked standard errors, see Green and Gerber 2012, p. 77.

³This design created 18 distinct block-types. We eventually used each type at least once. The data are made up of 505 blocks with two respondents in each block. Since respondents came through the project over a period of 12 days, blocking was done in real time. For example, if a 40 year old, white, female walked in and there was not an open block of that type, we started a new block for her. Our randomization routine assigned her to a mode even though her partner in the block had not yet appeared. This block remained open until the next 40 year old, white, female (or closest match within the categories) walked in, at which point she was automatically assigned to the opposite condition of the first respondent who opened the block; then the block was closed. Each block contains one respondent assigned to face-to-face and one respondent assigned to self-complete even though they might not have walked in to the experiment at the same time. We tracked block types and open blocks in real time.

⁴A table classifying all questions can be found in the appendix. A complete list of questions and wordings is available in the supplemental materials.

cover a variety of configurations with respect to non-response and the construction of question stems and outcome categories. There were five different types of non-response comparisons across the modes, as detailed below.

- **Explicit “don’t know” outcome.** In many cases, a don’t know outcome was available in writing in the self-complete survey and was offered out loud by the interviewer in the face-to-face survey. This describes 24 of the questions on the survey and offers a direct means of comparison with respect to non-response.
- **Voluntary “don’t know” vs. skipping.** In some cases (17 questions), we allowed interviewers to record when respondents voluntarily offered a don’t know response. This is a frequent setup for the ANES in-person survey. In the self-complete survey, however, volunteering an outcome was impossible. Respondents were told at the start of the survey that they could skip any questions they did not want to or could not answer.
- **No record of “don’t knows”.** For five of the questions we did not allow interviewers to record the don’t know outcome if a respondent volunteered it (if it was uttered, they recorded it as “Refused/Made other comment”). Similarly, we did not offer it as a response outcome in the self-complete survey (these respondents would have to skip the question).
- **Open-ended non-response.** The survey included three open-ended questions on which respondents could say whatever they liked. Interviews recorded exactly what respondents said in the in-person survey and respondents typed whatever they wanted in the self-complete survey. Respondents could type or say don’t know explicitly or they could leave the question blank or say nothing.
- **Voluntary “don’t know” vs. explicit “don’t know.”** In two cases, we allowed interviews to record voluntary don’t know responses in the face-to-face interview while offering the category explicitly in the self-complete survey. This combination

of outcomes is well-documented and produces a well known response pattern (more people will choose the outcome if it is offered as an option than otherwise, regardless of mode), thus we did not apply this scenario to many questions (see Mondak and Davis 2002; Mondak 2001; Delli Carpini and Keeter 1996).

Figure 1 includes all of the possible ways to provide a non-response across these different comparison sets (saying don't know, skipping, saying something else to the interviewer, etc . . .) on questions for which an explicit don't know option is offered in both modes. These questions have an explicit non-response choice written in the self-complete mode or the interviewer reads aloud a non-response option. Figure 2 includes all the other relevant questions in the survey (voluntary, open-ended, and other forms of non-response that are different across the modes). The most noticeable trend in Figures 1 and 2 is that more non-response occurs in the face-to-face interviews than in the self-completed surveys. Using a difference of means test and 95% confidence level, we are able to reject the null hypothesis that there is no difference between the modes for two-thirds of the 51 questions in the survey.

Figure 2 displays other types of comparisons. Separating the questions by non-response types shows similar proportions of questions with higher levels of non-response in the face-to-face interviews. For the questions with explicit don't know outcomes, 16 of 24 show significantly more non-response in the in-person interview relative to self-completed surveys; and for questions without explicit don't know outcomes, 11 of 17 questions show a similar pattern. This basic overview of the experimental findings suggests that there are differences in the rates of non-response across survey modes and more specifically, that there is something about the in-person interview that increases respondents' likelihoods of saying they do not know the answers to questions. Before we explore mechanisms for the potentially non-response-inducing effects of having an interviewer in the room, however, we consider other plausible explanations of this basic pattern of results that do not involve the presence of an interviewer.

Competing Explanations: Recency and Primacy

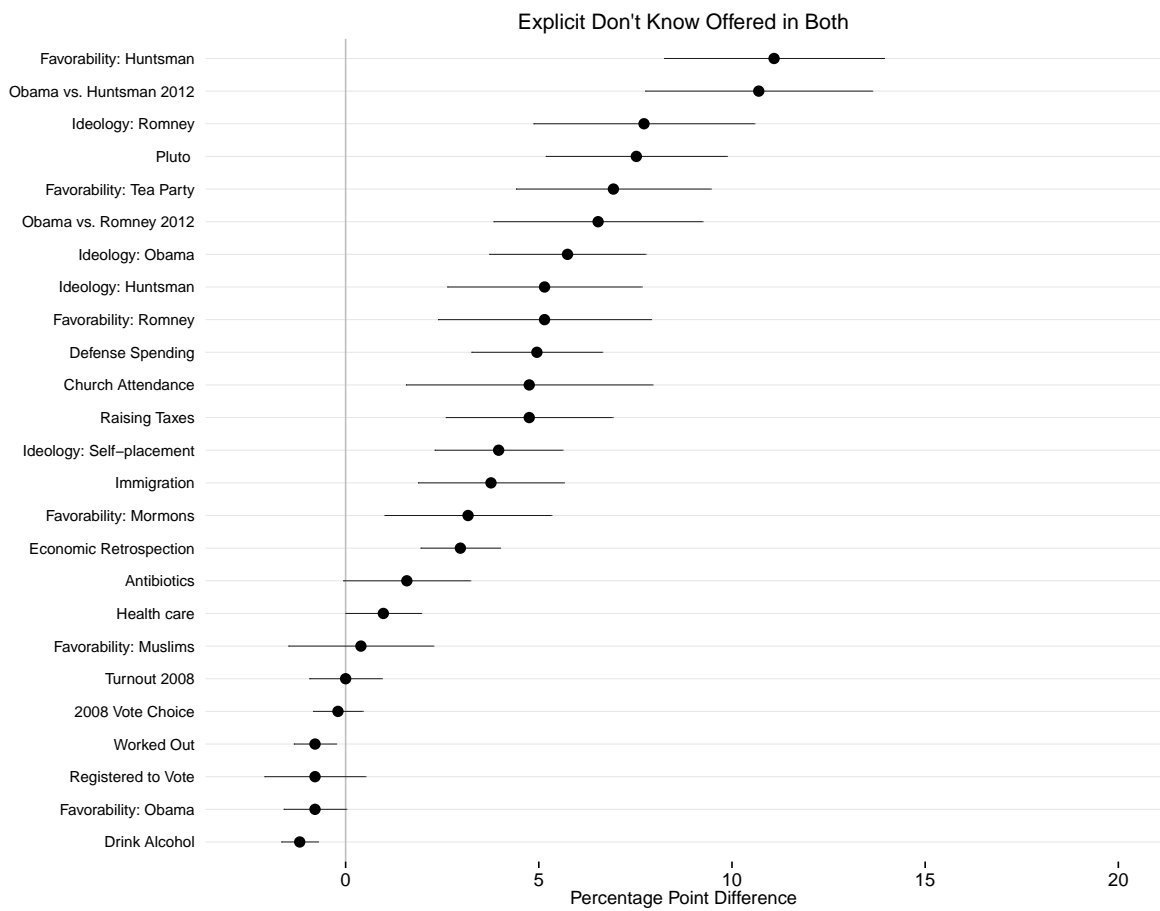
The fact that two-thirds of *all* types of questions show increased non-response in the face-to-face interview and that the magnitude of the differences in both figures is essentially the same helps to eliminate competing explanations for the pattern, such as a recency effect in the in-person interview (choosing the last outcome an interview read out loud) or a primacy effect in the self-completed surveys (choosing the first outcome listed under the stem). For example, if respondents in the in-person interview, on average, were choosing don't know options more frequently than respondents in the self-completed surveys because the words "or haven't you thought much about this?" were the last thing the interviewer said before pausing for an answer (recency), we would expect to see more and **larger** differences across the modes for questions that follow this form (Figure 1) relative to questions that did not follow this form (Figure 2). We see neither of these things in Figures 1 and 2.

The best evidence against recency as a driver of this effect is displayed by the solid, round plotting symbols in Figure 2, which show increased levels of *voluntary* non-response in the in-person interview relative to skipping in the self-completed surveys. That is, more people in the face-to-face group told the interviewer they did not know the answer to a question *even when it was not mentioned as a possible outcome in the stem of the question* than skipped that same question in the self-completed survey. This pattern is made more striking by the realization that in this comparison the question wording is *exactly* the same across the two modes. The only difference is that respondents can tell an interviewer they do not know the answer whereas they have to simply skip the question to indicate they do not know the answer in the self-complete mode (which they have previously been told they can do with no penalty). This pattern, and the similar size and frequency of increased non-response in face-to-face interviews across question types with and without don't know outcomes as the final choice, eliminates recency as a plausible driver of non-response. In both cases, when don't know is offered by interviewers (and on screen) as the

final outcome and when it is not offered at all, there is more non-response, on average, in the face-to-face mode.

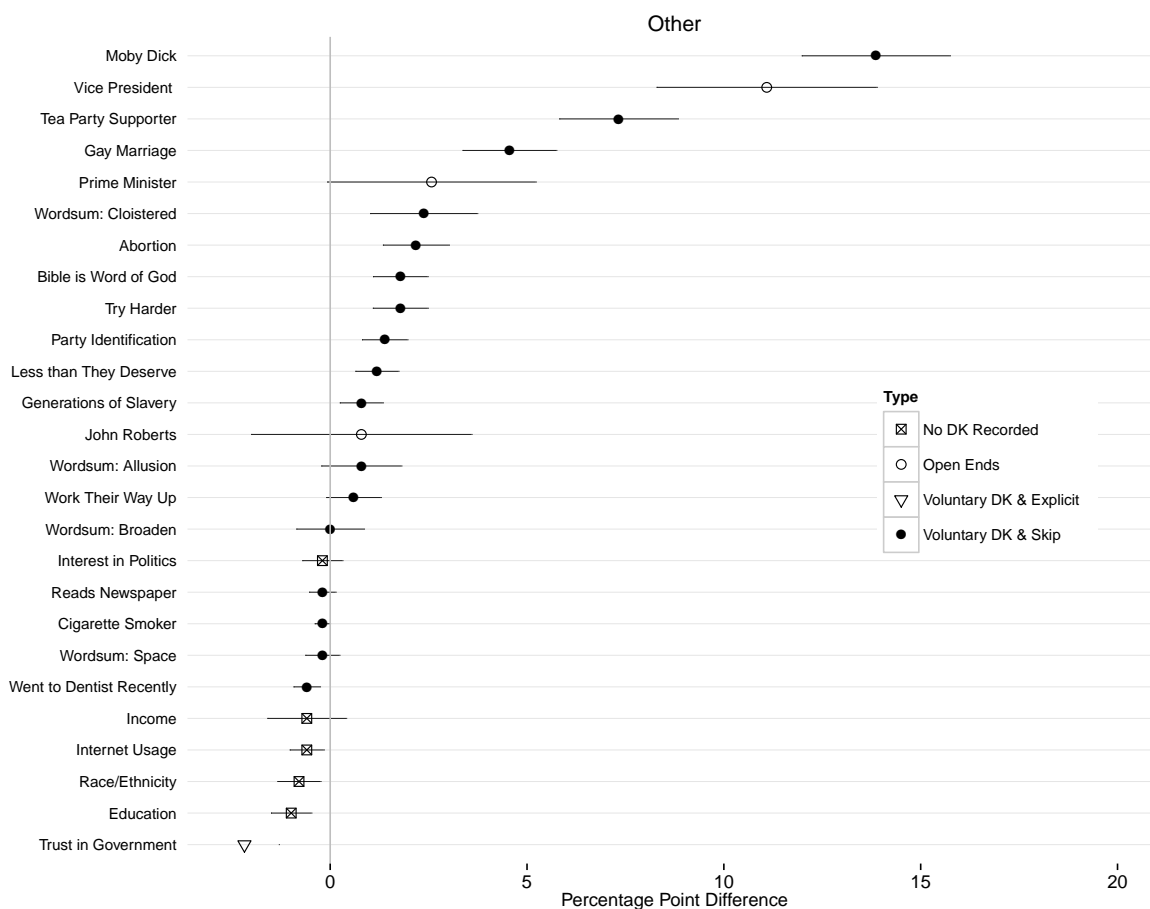
Similarly, tests for primacy effects (choosing the first outcome listed on the screen in the self-complete survey) suggest that, within questions, respondents are not choosing the primary outcome more than other outcomes, nor are they choosing the primary outcome in greater magnitude in the self-complete than in the face-to-face interview (see Bishop and Smith (2001) for corroborating experimental evidence that visual versus aural modes of survey administration do not produce different rates of recency or primacy effects and that when these effects exist they are quite small).

Figure 1: Non-response by Mode for Questions with Explicit Don't Know



Note: Percent of non-response in face-to-face minus self-complete surveys, such that positive values indicate more non-response in the face-to-face survey, with 95% confidence intervals around each difference.

Figure 2: Non-response by Mode for All Other Questions



Note: Percent of non-response in face-to-face minus self-complete surveys, such that positive values indicate more non-response in the face-to-face survey, with 95% confidence intervals around each difference.

Non-response by Topic: Patterns and Mechanisms

Having ruled out basic survey design factors like recency and primacy as drivers of the patterns displayed in Figures 1 and 2, we move to a more nuanced assessment of the differences in non-response across mode by examining the patterns across different types of questions. This analysis provides leverage in identifying possible mechanisms that involve the presence of an interviewer for the patterns in these figures. The first set of results to provide some insight into what might be going on is that questions about attitudes,

evaluations, issue positions, and factual knowledge exhibit increased non-response in the face-to-face mode, but reports of behavior and demographic indicators show no such increase.

A rich literature on the psychology of survey response helps in understanding this pattern. The primary pattern, that behaviors show no differences relative to attitudes, is explained nicely with theories of survey response, which suggest that people remember things they experience first-hand better than things they learn about second-hand. Activities and descriptors like voting in elections, going to the dentist, annual income, smoking cigarettes, or reading the newspaper are part of what social psychologists call an “ongoing autobiographical narrative” and thus memory of such things, and survey responses drawing on those memories, are aided by things like sensory detail and other psychological processes (Tourangeau, Rips, and Rasinski 2000, p. 65). The answers to these questions are “real” to respondents and thus easy to recall. Remembering and reporting on these autobiographical details is routine for most people, thereby increasing the stability of responses to these types of questions across surveys over time. By extending this line of reasoning from within to across subjects, we believe the same mechanism explains the comparability, on average, of behavioral reports across our two modes of interview: first-hand experiences are easy for people to recall and thus other factors that usually make answering survey questions difficult are less relevant, regardless of mode.

Responses to questions about attitudes and beliefs, on the other hand, are quite different from first-hand memories. Tourangeau and his colleagues (2000) write that, “attitude questions pose special problems” for respondents because people may never have considered the topic in question prior to hearing the survey question (p. 62). Increased response instability over time (and we suggest mode of inquiry, too) is one result of the higher level of cognitive burden associated with questions about attitudes and beliefs, relative to those about behaviors and demographics. Tourangeau, Rips, and Rasinski (2000, p. 179) detail the mechanisms from which the instability arises in their belief-sampling model of survey

response. The model traces response differences on attitude questions to variations in respondents' retrieval and judgment processes across survey settings. Which considerations a respondent retrieves and places weight on depends on the momentary accessibility of each consideration, which according to the model, is influenced by many factors, some chronic and some temporary. The accessibility of such considerations, it seems, can also be context — or mode — dependent. Tourangeau, Rips, and Rasinski (2000) consider this exact possibility, writing that judgments about considerations may also be affected by momentary fluctuations, including things like the presence of an interviewer (p. 180).

In summary, differences across mode in responses to questions about attitudes, beliefs, and knowledge, may differ depending on the extent to which respondents consider the same information in both modes and construe that information in the same way across modes. In this experiment, the differences in context include the presence of the interviewer in the face-to face setting, the need to process information either aurally or visually, and the required use of a keyboard in the self-completed mode. Any or all of these factors may influence the momentary accessibility of considerations and affect the retrieval and judgment processes respondents bring to bear on answering attitude-based questions across the different modes of interview.

Specific Questions about Attitudes, Evaluations, & Beliefs

In the survey, questions drawing on attitudes and beliefs consist of evaluations of candidates and groups, hypothetical presidential elections match-ups, and questions about political issues. We asked respondents to rate President Obama on favorability and ideology, along with the most well-known Republican party contender at the time, Mitt Romney, and the least well-known of the GOP field, Jon Huntsman (according to national polls in the field at the time). Among attitude-based questions, we find the biggest difference in non-response between the modes on questions involving Huntsman, which

is expected since most people knew very little about him and were therefore less likely to have well-formed considerations about him. In the face-to-face interview nearly 75% of respondents failed to rate Huntsman on favorability. In the self-complete mode, fewer people (63%) opted out of rating Huntsman – a drop of 11 points. Similarly, when we asked respondents who they would vote for in November if the election were a choice between Barack Obama and Jon Huntsman, 43% said they didn't know when an interviewer asked, but only 33% failed to choose in the self-complete survey.

Differences in the same direction persist for questions about perceived ideological placements of candidates and for ideological self-placements as well, although they are not quite as large as for favorability ratings (ranging from four to seven points). In terms of issue positions, we asked respondents about their positions on the following issues in order of declining levels of differences in non-response: defense spending, raising taxes, gay marriage, immigration, the state of the economy, abortion, and healthcare. All of the issues exhibit a statistically significant higher rate of non-response in the face-to-face survey, although the differences for abortion and health care are substantively small. These findings demonstrate that respondents are more likely to answer “I don't know” to attitude-based questions when the survey is administered by an interviewer in the room with them when compared to a survey completed on a computer.

In determining what to make of this pattern, it is important to bear in mind that while we typically think of non-response as a normatively bad thing, there is no reason to assume that the increased level of non-response in the in-person interview is an artifact of the mode while the self-completed results reflect respondents' true underlying opinions or knowledge with more fidelity. Since all of the attitude questions are subjective in nature, we have no way of ascertaining which mode comes closer to “the truth.” Anticipated this situation, we designed the survey with a set of questions to help us sort out which mode better reflected what respondents know or believe they know.

We asked six questions with factually correct answers. To the extent that we see

increased levels of correct responses across the questions in the self-complete mode coinciding with lower levels of non-response, we will have some purchase on whether the interviewer is somehow influencing which considerations a respondent brings to bear on answering a question (as described earlier), in this case, potentially by interacting with respondent characteristics, exacerbating things like confidence, anxiety, or the willingness to take risks — and ultimately garnering a non-response from someone in the in-person setting who might actually have a substantive answer to the question that they choose not to give.

Questions about Facts

Decades of research in education testing demonstrates that when students are offered the chance to say “don’t know” on test questions, knowledge assessments suffer because the questions tap two different constructs: knowledge and the propensity to make a guess. This propensity, not surprisingly, varies with personality traits like self-confidence (Casey, Nuttall, and Pezaris 1997; Hirschfeld, Moore, and Brown 1995), likelihood of taking risks (Ben-Shakhar and Sinai 1991; Cronbach 1946), and competitiveness (Hirschfeld, Moore, and Brown 1995). Cronbach (1946, p. 491) explains that these different propensities are a “tendency” that leads a person to consistently “make different responses to test items than he would have made had the same content been presented in a different form.” Mondak and Davis (2001), using a split-half experimental design across three different interviewer-assisted surveys at different points in time, demonstrate that reducing the share of don’t know responses (by eliminating it as an outcome category) increased the number of correct responses overall and per capita (2001, p. 212). We demonstrate the same pattern here: for each of our fact-based questions, non-response goes down and correct answers go up, but not because of the elimination of don’t know as an outcome category. The cause in our experiment is the mode of survey administration: taking the interviewer out of the survey experience decreases non-response and increases the number

of correct answers on average and per capita. We present these differences in Figure 3 for open-ended political knowledge questions and in Figure 4 for non-political closed-ended questions.

We asked three canonical open-ended political knowledge questions about who holds the job of Vice President, Prime Minister of the U.K., and Chief Justice of the Supreme Court. In the first two cases, respondents had to come up with the name of the person who holds the job with no choice set provided and in the last case they had to name the job held by John Roberts.⁵ These questions resemble those used in the ANES, but we coded the answers with slightly more nuance than is usually provided by the ANES.⁶ The most striking difference comes from the question asking respondents to name the sitting Vice President of the United States. Roughly 61% of respondents, on average, got this question right and 55% got it exactly right. The difference by mode is striking: only 50% of respondents in the in-person interview came up with the name Joe Biden, but slightly more than 60% of self-complete respondents did so; and there is an increased rate of non-response in the face-to-face interview of more than 10 points.⁷

This pattern is repeated for the questions about John Roberts and the Prime Minister of the U.K., although in smaller magnitudes. Only 21% of respondents could identify

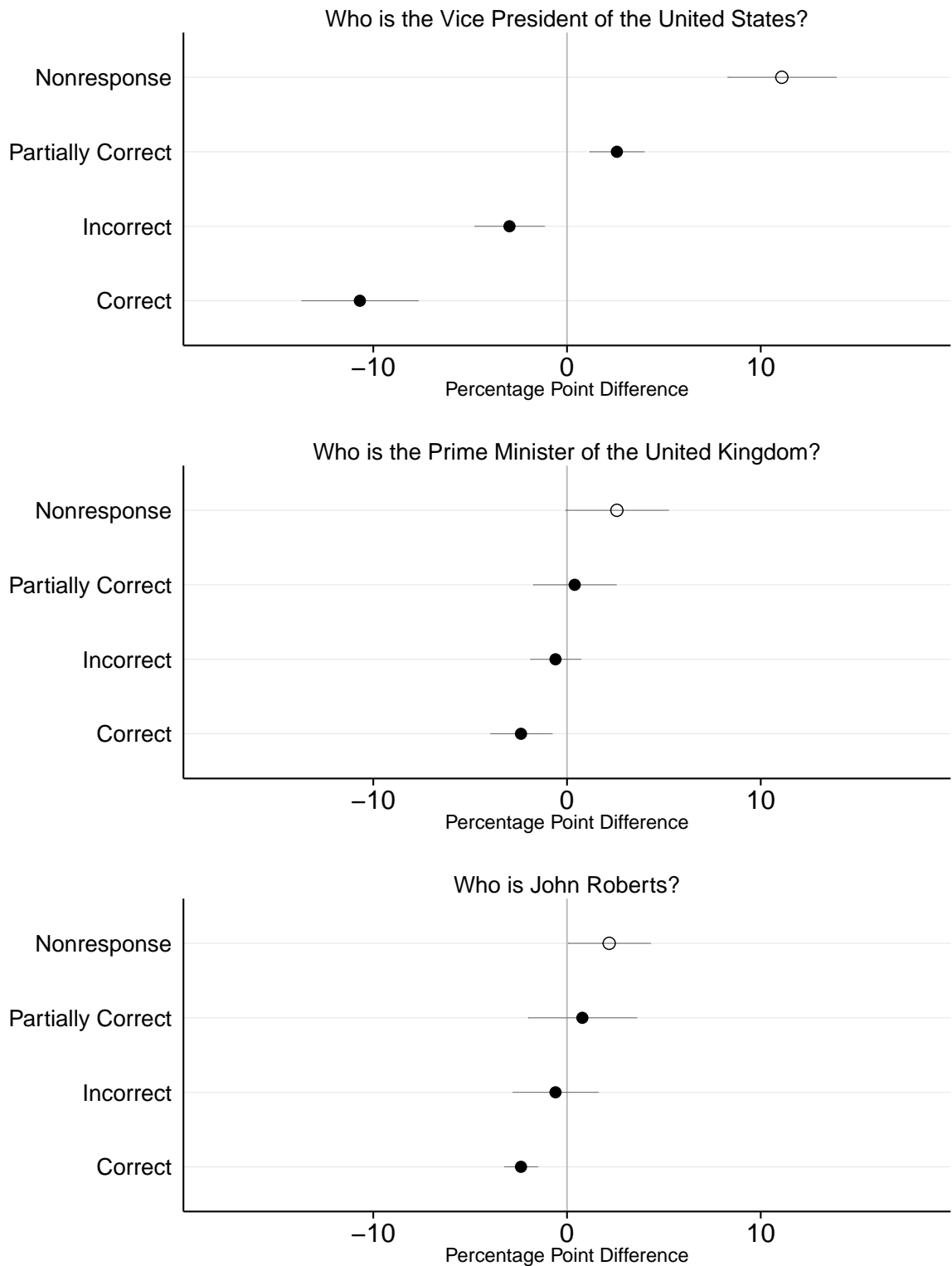
⁵One funny situation arose in an in-person interview when a respondent shouted out, “I’m a builder!” in response to the John Roberts item. His name, he explained, was John Roberts.

⁶Responses that were flatly incorrect were coded as such and separated from any kind of statement indicated non-response. We also gave credit for partial answers. For example, answers to the vice president question containing identifying comments about Joe Biden instead of naming him (like “the white haired guy from Scranton”) were coded as partially correct.

⁷To be sure that respondents in the computer surveys were not looking up the answers to the factual questions, we downloaded the browser history of each self-completed respondent upon completion of the interview to confirm that no one had looked anything up online. Even though Google, Wikipedia, and an endless number of reference sites were easily available, only two respondents cheated on the knowledge questions by looking up the answers. Unprompted, one respondent admitted to us after completing the survey that he looked up the name of the Vice President. These two respondents were coded as incorrect.

David Cameron as Prime Minister and even fewer, 16%, knew that John Roberts was the Chief Justice of the Supreme Court. The differences across modes for these questions tapping relative low levels of overall knowledge were not as big — about a three-point increase in non-response in the face-to-face mode and a similarly sized increase in correct answers in the self-complete.

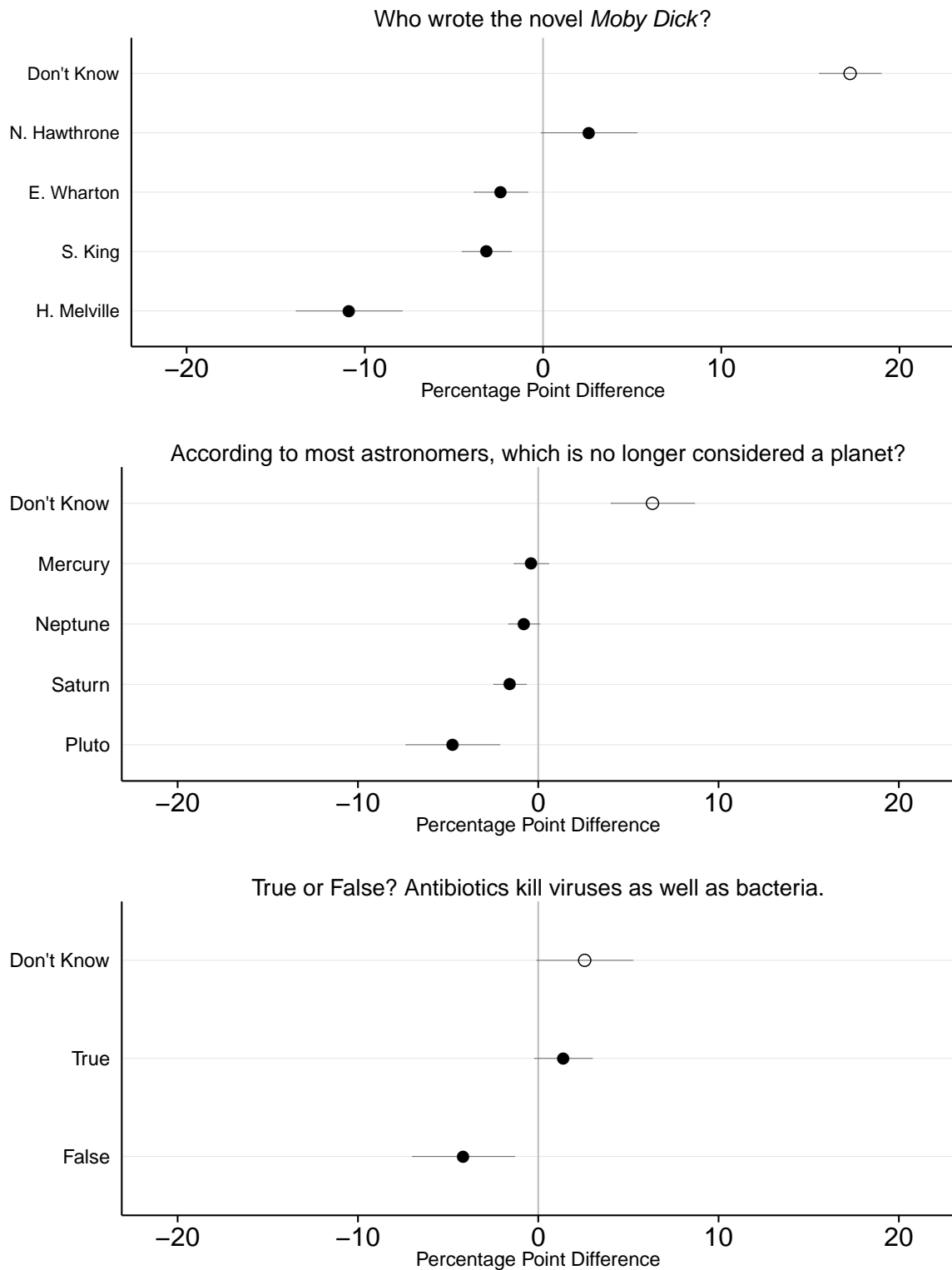
Figure 3: *Mode Differences in Open-ended Political Knowledge Questions*



Note: Plotting symbols reflect face-to-face values minus self-complete. Bars represent 95% confidence intervals. Hollow symbols indicate non-response.

We also asked three closed ended, non-political questions frequently used by Pew. These questions ask respondents to identify which celestial object astronomers no longer considered to be a planet in our solar system, name the author of *Moby Dick*, and choose whether it was true or false that antibiotics kill viruses as well as bacterial infections. We present the differences by mode in Figure 3. In each case, a familiar pattern emerges: More non-response in the face-to-face interview and increased rates of correct responses in the self-complete. The most striking difference comes from the question about *Moby Dick*, which 52% of respondents correctly answered. There is nearly a 20-point increase in non-response for this item in the face-to-face mode. The increase in the number of people saying Herman Melville was slightly greater than 10-points for the self-complete responses. Similarly, there is nearly a 10-point increase in non-response in the face-to-face mode on the Pluto question (74% of respondents knew this on average) and a five-point increase in non-response on the antibiotics question (66% knew this).

Figure 4: *Mode Differences in Non-response for Closed-ended Knowledge Questions*



Note: Plotting symbols reflect face-to-face values minus self-complete. Bars represent 95% confidence intervals. Hollow symbols indicate non-responses. For the *Moby Dick* question, a DK option was not explicitly given in either mode. Explicit DKs appeared in both modes for Pluto and Antibiotics.

Across questions about attitudes, beliefs, and facts, both closed and open-ended, about politics and non-political topics, the same pattern emerges: in-person interviewing generates higher levels of item non-response than self-completed surveys. Recency effects can be ruled out as a mechanism for these results as we observe the same pattern when no non-response outcome is read aloud to respondents at all (or offered in the self-complete survey); and primacy effects can be ruled out as we varied the order in which outcomes were offered in the self-completed surveys and find no confirmatory primacy pattern. Although we are not in a position to illustrate the exact cause of the difference in non-response by mode, the increased incidence of factually correct answers in the self-completed surveys relative to the in-person interviews mimics the findings in educational testing (Casey, Nuttall, and Pezaris 1997; Hirschfeld, Moore, and Brown 1995; Ben-Shakhar and Sinai 1991; Cronbach 1946), and by Mondak and Davis (2001) in political science specifically, demonstrating the effects of encouraging people to say “I don’t know.”

Our experimental results taken in context with previous findings suggest that when the interviewer is removed from the survey experience, respondents’ personal characteristics interact less significantly with the administration of the survey. To investigate this, we explore variation on respondent characteristics and their relationship to increased levels of non-response in face-to-face surveys.

Modeling Non-response

We begin with a simple, individual-level model of item non-response. There are many known predictors of item non-response (Heerwegh and Loosveldt 2008; Berinsky 2004; Althaus 2003; Holbrook, Green, and Krosnick 2003; Tourangeau, Rips, Rasinski 2000; Delli Carpini and Keeter 1996; Krosnick 1991; Rapoport 1982; Converse 1976; Francis and Busch 1975; Converse 1976; Converse 1964; 1970) and we will control for as many of them as possible before focusing on two things in particular – a respondent’s level of

cognitive skill and its interaction with survey mode. In their student-based experiment, Chang and Krosnick (2010) found that college-reported standardized test scores (like SAT scores from the admissions office) interacted with the mode of interview such that students with lower levels of aptitude (as predicted by their SATs) demonstrated lower levels of concurrent validity and non-differentiation in the interviewer-assisted mode than in the self-administered mode. We build on that finding by examining the role of cognitive skill in an adult population and extending the analysis to item non-response. If these data corroborate Chang and Krosnick's result among adults, with a real face-to-face interview and a sub-set of proven measures of cognitive skill from the General Social Survey (GSS), we will have learned something specific about how the mode of interview is affecting particular respondents with respect to non-response.

We begin by summing item non-response within the six categories of questions in the survey: issue positions, behaviors, ideology, favorability ratings, vote choice, and facts. The questions that form each index are summarized in Table 1. The total range of each index is on the bottom row of Table 1.

Table 1: Questions included in Non-Response Index by Topic

Issues	Behavior	Ideology	Favorability	2012 Vote	Facts
Taxation	2008 Vote	Obama	Obama	Obama-Huntsman	Moby Dick
Defense	Turnout	Huntsman	Huntsman	Obama-Romney	Pluto
Gay Marriage	Register	Romney	Romney		Antibiotics
Immigration	Drink	Self	Muslims		
Health care	Gym		Mormons		
Abortion	Church		Tea Party		
Internet Usage					
0-6	0-7	0-4	0-6	0-2	0-3

Note: Items summed over non-response.

Our models of non-response include indicators for African American and Hispanic ethnicity, education, income, three levels of age, party identification, three levels of political knowledge, gender, and mode of interview. As a measure of cognitive skill, we use four of the ten WORDSUM vocabulary items that have appeared regularly on the GSS since 1972, choosing items from each of the difficulty levels known to provide discrimination. The WORDSUM items ask respondents to choose the closest synonym from among 5 choices for the word presented in the question stem. Respondents are handed a show-card with the word in question typed in bold at the top and the choices in regular type underneath. The interviewer also reads the word and the choices out loud to the respondent. These items have been shown repeatedly to be good indicators of cognitive skill and general intelligence (Davis, Smith and Marsden 2007; Carroll 2003; Jensen 2001; Thorndike 1942).⁸ We are not only interested in whether this scaled measure of apti-

⁸We scale these four items using a two-parameter item response model and calculate

tude affects item non-response across our topics, but also whether it interacts with the mode of interview in a particular way. Specifically, we have in mind Chang and Krosnick's (2010) claim that low-skilled respondents have a harder time answering questions in interviewer-administered surveys than they do in self-administered ones.

We present the results of these investigations in Table 2.⁹ As expected, the indicator for face-to-face mode assignment shows a significant increase in non-response across all the topics except reports of behavior (as discussed previously). To get a sense of the magnitude of the effects, consider the results for questions about political ideology. We asked respondents to place Romney, Obama, Huntsman, and themselves on a five-point ideology scale. Non-response for this set of questions can range from 0 to 4. The average difference between the modes of interview on ideology is a .19 increase in the in-person interview. This is roughly a 4 percent increase in non-response due purely to mode. Similarly, the data reveal a 5 percent increase in non-response in the in-person interview on questions of fact, and a 7 percent increase for questions about vote choice.

Even controlling for education and level of political information, cognitive skill has a direct effect on non-response for half of the topics we investigate (increasing skill leads to less non-response, all else equal). In most cases, the size of the effect is considerably smaller in magnitude than the effect due to mode of interview. To illustrate the consequences of face-to-face interviews and put these results in context, consider that for questions about issue positions we would have to make people one standard deviation smarter to produce the same size effect that is obtained from shifting to a self-complete instrument.

An important body of research on public opinion was born from the result (Converse

a latent ability score for each respondent. The range of the scale is constructed to have a mean of 0 and range from 2 to -2. Higher values indicate lower levels of cognitive skill such that the interaction with mode of interview produces *increases* in non-response. See Malhotra et al. (2007) for discussion of using a short version of the WORDSUM scale.

⁹The control indicators are not remarkable — education, partisanship, political knowledge, and gender are generally predictors of non-response across the various topic areas with the largest decreases in non-response coming from increasing education, being a partisan, and high levels of political information, all else equal.

Table 2: Models of Non-response by Question Topic

Covariates	Determinants of Non-responses					
	Issues	Behavior	Ideology	Favorability	2012 Vote Choice	Facts
(intercept)	.68 (.10)	.42 (.09)	1.89 (.13)	1.96 (.16)	1.12 (.11)	.08 (.08)
Face-to-face	.18 (.05)	.01 (.04)	.19 (.06)	.22 (.07)	.15 (.05)	.22 (.04)
Cognitive Skill (\uparrow = Lower Cog. Skill)	.11 (.03)	.08 (.03)	.05 (.04)	.02 (.05)	.02 (.04)	.09 (.03)
Cognitive Skill * Face-to-face	-.01 (.05)	.06 (.04)	.10 (.06)	.15 (.08)	.10 (.05)	.11 (.04)
African American	.06 (.07)	.05 (.03)	.08 (.08)	.13 (.11)	-.36 (.07)	.17 (.05)
Hispanic	.14 (.07)	.00 (.06)	.19 (.09)	.36 (.12)	-.02 (.08)	.03 (.06)
High School	.17 (.06)	.22 (.05)	.33 (.07)	.26 (.10)	.03 (.06)	.12 (.05)
Some College or 2 yr. degree	.03 (.04)	.03 (.04)	-.09 (.05)	.08 (.07)	.04 (.04)	-.01 (.03)
College	.02 (.04)	.03 (.04)	-.12 (.05)	-.14 (.07)	.07 (.05)	-.01 (.03)
Graduate School	-.01 (.05)	-.07 (.04)	-.25 (.06)	-.23 (.08)	-.10 (.05)	-.10 (.04)
Income	-.02 (.01)	-.01 (.00)	-.01 (.00)	-.01 (.01)	.00 (.01)	-.01 (.01)
Age	.00 (.00)	.00 (.00)	-.01 (.00)	-.01 (.00)	-.01 (.00)	.01 (.00)
Female	.12 (.05)	.06 (.04)	.29 (.06)	.42 (.08)	.06 (.05)	.07 (.04)
Partisan	-.19 (.05)	.03 (.04)	-.09 (.06)	-.09 (.08)	-.35 (.05)	-.02 (.04)
Political Knowledge: middle	-.04 (.06)	-.07 (.05)	-.44 (.08)	-.47 (.10)	-.26 (.07)	-.06 (.05)
Political Knowledge: high	-.05 (.11)	-.18 (.09)	-.97 (.13)	-.86 (.17)	-.18 (.11)	-.12 (.08)
R-squared	.12	.09	.24	.20	.14	.16
Sample size	1010	1010	1010	1010	1010	1010

Note: Cell entries are OLS coefficients with standard errors in parentheses. The dependent variable in each model is the total number of non-responses within topic (see Table 1).

1964) that people cannot place candidates on ideological and issue scales. These results raise the possibility that some of these conclusions may have been artifacts of the mode of interview.

Additional and heterogeneous costs of the in-person interview are evident when we examine the interaction between survey mode and cognitive abilities. The effects of face-to-face interviewing are significantly greater for respondents with low levels of skill. For three of the six topic areas, the interaction between mode of interview and cognitive skill is both substantively important and statistically different from zero — that is, for most of the questions on the survey, not only was there an increase in non-response in face-to-face interviews relative to self-complete interviews, but this increase is exacerbated for respondents with low levels of cognitive skill.¹⁰ For questions about favorability, vote choice, and factual knowledge, the difference in face to face interviews for people with low levels of cognitive skill compared to those with average levels is a 36 to 63% increase in non-response. To demonstrate the way cognitive skill interacts with mode of an interview, consider two otherwise average respondents one with the lowest level of cognitive skill the other with the highest. The difference in the amount of non-response between these two respondents is minimal for the self-complete mode of interview — moving from the highest skill level to the lowest, all else equal, increases non-response by roughly 6%. But in the in-person mode, moving from high skills to low increases non-response by 35%. That is a nearly 500% increase in the rate of skill-based changes in non-response due to mode of interview alone.

¹⁰For questions about ideology the interaction is substantively similar to these other areas but with a lower level of statistical precision.

Conclusion and Implications

The results of this randomized trial assigning respondents to different modes of survey after they agree to participate suggest that the presence of an interviewer affects the answers people give, particularly with regard to non-response. Nearly two-thirds of the questions on the survey we administered exhibited significantly higher levels of non-response from respondents assigned to the face-to-face survey compared to those assigned to the self-complete survey. These patterns exist across multiple styles and formats of questions with varying ways of expressing non-response and they do not seem to be driven by primacy or recency of outcomes.

More importantly, however, our findings show that certain kinds of people are more affected by the presence of an interviewer than others. The in-person interview is more likely to lead to item non-response for respondents with low levels of cognitive abilities relative to those with high levels. These results obtain even controlling for respondent's levels of education and political knowledge. While this experiment is the first to demonstrate this pattern on an adult population, it is the second to demonstrate the general link between cognitive ability and in-person interviewing (see Chang and Krosnick 2010). The increased chances of observing a non-response by low-skilled individuals in face-to-face interviewing should lead not only to a reconsideration of public opinion findings about how little people know about politics, but also to the belief that in-person interviewing is the gold standard on which high quality surveys must rely. A very particular group of people in the population are offering "don't know" responses to an interviewer when their counterparts in self-completed surveys are, for example, choosing correct answers to purely fact-based questions.

As the costs of conducting in-person interviews for large national projects like the ANES escalate and the demands for more statistical power and design-based inference increase, it is worth considering whether the in-person interview is worth the substantial

investment it demands. A switch to self-completed surveys cannot be far off given the current funding climate, but such a switch need not be viewed as a negative compromise. To the contrary, it seems such a move may increase the quality of the data we rely upon to characterize what people know and think about politics.

References

- Althaus, Scott L. 2003. *Collective Preferences in Democratic Politics: Opinion Surveys and the Will of the People*. New York: Cambridge University Press.
- Acree, Michael, Maria Ekstrand, Thomas J. Coates and Ron Stall. 1999. "Mode Effects in Surveys of Gay Men: A Within-individual Comparison of Responses by Mail and Telephone." *The Journal of Sex Research* 36 (1): 67-75.
- Berinsky, Adam J. 2004. *Silent Voices: Public Opinion and Political Participation in America*. Princeton, NJ: Princeton University Press.
- Ben-Shakhar, Gershon, and Yakov Sinai. 1991. "Gender Differences in Multiple-Choice Tests: The Role of Differential Guessing Tendencies." *Journal of Educational Measurement* 28 (1): 23-35.
- Bishop, George F., Hans-Juergen Hippler, Norbert Schwarz and Fritz Strack. 1988. "A Comparison of Response Effects in Self-Administered and Telephone Surveys." In *Telephone Survey Methodology*, eds. Robert M Groves, Paul P. Biemer, Lars E. Lyberg, James T. Massey, William L. Nicholls and Joseph Waksberg. New York: Wiley Press.
- Bogart, Leo. 1972. *Silent Politics: Polls and the Awareness of Public Opinion*. New York: Wiley-Interscience Press.
- Bradburn, Norman M. and Seymour Sudman. 1988. *Polls and Surveys: Understanding What They Tell Us*. San Francisco: Jossey-Bass Press.
- Chang, Linchiat and Jon A. Krosnick. 2010. "Comparing Oral Interviewing with Self-Administered Computerized Questionnaires: An Experiment." *Public Opinion Quarterly* 74 (1): 154-67.

- Carroll, John B. 2003. "The Higher-Stratum Structure of Cognitive Abilities: Current Evidence Supports G and about Ten Broad Factors." In *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, eds. Helmuth Nyborg. New York: Pergamon Press.
- Casey, M. Beth, Ronald L. Nuttall, and Elizabeth Pezaris. 1997 "Mediators of Gender Differences in Mathematics College Entrance Test Scores: A Comparison of Spatial Skills with Internalized Beliefs and Anxieties." *Developmental Psychology* 33 (4): 669-80.
- Coombs, Clyde H. and Lolagene C. Coombs. 1976. "'Don't Know': Item Ambiguity or Respondent Uncertainty?." *Public Opinion Quarterly* 40 (4): 497-514.
- Converse, Jean M. 1976. "Predicting No Opinion in Polls." *Public Opinion Quarterly* 40 (4): 515-30.
- Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, eds. David E. Apter. New York: Free Press.
- Converse, Philip E. 1970. "Attitudes and Non-Attitudes: Continuation of a Dialogue." In *The Quantitative Analysis of Social Problems*, eds. Edward R. Tufte. Reading: Addison-Wesley Press.
- Cronbach, Lee J. 1946. "Response Sets and Test Validity." *Educational and Psychological Measurement* 6 (4): 475-94.
- Davis, James A., Tom W Smith and Peter V Marsden. 2007. "General Social Survey, 1972-2006: Cumulative Codebook." Chicago: National Opinion Research Center.
- de Leeuw, Edith D. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: T.T.-publikaties.
- Delli Carpini, Michael X. and Scott Keeter. 1996. *What Americans Know about Politics and Why it Matters*. New Haven, CT: Yale University Press.

- Feick, Lawrence F. 1989. "Latent Class Analysis of Survey Questions That Include Don't Know Responses." *Public Opinion Quarterly* 53 (4): 525-47.
- Fowler, Floyd J., Anthony M. Roman, Zhu Xiao Di. 1998. "Mode Effects in a Survey of Medicare Prostate Surgery Patients." *Public Opinion Quarterly* 62 (1): 29-46.
- Francis, Joe D. and Lawrence Busch. 1975. "What We Now Know about 'I Don't Know'." *Public Opinion Quarterly* 39 (2): 207-18.
- Gano-Phillips, Susan and Frank D. Fincham. 1992. "Assessing Marriage via Telephone Interviews and Written Questionnaires: A Methodological Note." *Journal of Marriage and Family* 54 (3): 630-35.
- Heerwegh, Dirk and Geert Loosveldt. 2008. "Face-to-face Versus Web Surveying in a High-Internet-Coverage Population: Differences in Response Quality." *Public Opinion Quarterly* 72 (5): 836-46.
- Heerwegh, Dirk. 2008. "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 (1): 111-121.
- Hirschfeld, Mary, Robert L. Moore, and Eleanor Brown. 1995. "Exploring the Gender Gap on the GRE Subject Test in Economics." *Journal of Economic Education* 26 (1): 3-15.
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Versus Face-to-face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias." *Public Opinion Quarterly* 67 (1): 79-125.
- Jensen, Arthur R. 2001. "Vocabulary and General Intelligence." *Behavioral and Brain Science* 24 (6): 1109-10.

- Kiesler, Sara and Lee S. Sproull. 1986. "Response Effects in the Electronic Survey." *Public Opinion Quarterly* 50 (3): 402-13.
- Kinder, Donald R. and Lynn M. Sanders. 1996. "Divided by Color: Racial Politics and Democratic Ideals." Chicago: University of Chicago Press.
- Krosnick, Jon A. 2010. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (3): 213-36.
- Krosnick, Jon A, Allyson L. Holbrook, Matthew K. Berent, Richard T. Carson, W. Michael Hanemann, Raymond J. Kopp, Robert Cameron Mitchell, Stanley Presser, Paul A. Ruud, V. Kerry Smith, Wendy R. Moody, Melanie C. Green, and Michael Conaway. "The Impact of 'No Opinion' Response Options on Data Quality: Non-attitude Reduction or an Invitation to Satisfice?." *Public Opinion Quarterly* 66 (3): 371-403.
- Krosnick, Jon A. and Arthur Lupia. 2008. "Problems with ANES Questions Measuring Political Knowledge." Stanford University. Unpublished paper.
- Oppenheim, Abraham N. 1992. *Questionnaire Design, Interviewing, and Attitude Measurement*. London: Pinter Press.
- Malhotra, Neil, Jon A. Krosnick and Edward Haertel. 2007. "The Psychometric Properties of the GSS Wordsum Vocabulary Test." Chicago: National Opinion Research Center.
- Malhotra, Neil, and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples." *Political Analysis* 15 (3): 286-323.

- Mondak, Jeffery J., and Belinda Creel Davis. 2001. "Asked and Answered: Knowledge Levels When We Will Not Take 'Don't Know' for an Answer." *Political Behavior* 23 (3): 199-224.
- Mondak, Jeffery J. 2001. "Developing Valid Knowledge Scales." *American Journal of Political Science* 45 (1): 224-238.
- Noelle-Neumann, Elizabeth. 1993. *The Spiral of Silence: Public Opinion, Our Social Skin*. Chicago: University of Chicago Press.
- Rapoport, Ronald B. 1983. "Sex Differences in Attitude Expression: A Generational Explanation." *Public Opinion Quarterly* 46 (1): 86-96.
- Sanders, David, Harold D. Clarke, Harold D., Stewart, Marianne C. and Paul Whiteley "Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study." *Political Analysis* 15 (3): 257-285.
- Schwarz, Norbert and Gerd Bohner. 2001. "The Construction of Attitudes." In *Blackwell Handbook of Social Psychology: Intraindividual Processes*, eds. Abraham Tesser and Norbert Schwarz. New York: Wiley Press.
- Sears, David O. 1986. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature." *Journal of Personality and Social Psychology* 51 (3): 515-30.
- Strack, Fritz, Norbert Schwarz, and Michaela Wanke. 2010. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5 (3): 213-36.
- Sudman, Seymour and Norman M. Bradburn. 1974. *Response Effects in the Electronic Survey*. Chicago: Aldine Press.
- Thorndike, Robert L. 1942. "Two Screening Tests of Verbal Intelligence." *Journal of Applied Psychology* 26 (2): 128-35.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. New York: Cambridge University Press.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.