American National Election Studies (ANES)
http://www.electionstudies.org

Board Report

Assessing the Implications of a Mode Change

D. Sunshine Hillygus (Duke University), chair
Nicholas Valentino (University of Michigan)
Lynn Vavreck (University of California, Los Angeles)
Matt Barretto (University of California, Los Angeles)
Geoff Layman (Notre Dame)

May 14, 2017

Assessing the Implications of a Mode Change

May 14, 2017

The Board of the American National Election Study (ANES) established a set of committees to assess the future of the project. The ANES has been the "gold standard" election survey since its inception in 1948. It is now the longest-running collection of American political attitudes in existence. Rising survey costs, budgetary concerns, and technological changes are creating pressure to evaluate less expensive design options, including and especially a potential shift from the face-to-face interviews to a self-completed survey conducted online. This report presents the conclusions of the committee charged with assessing the implications of such a mode change. In assessing this change, we have placed a premium on the continuity of the ANES time-series, although other considerations are also important (for example, differences across the modes with respect to costs, bias, and precision of estimates). In privileging continuity, we will most often be comparing the modes to one another. Our view is that it is not this committee's charge to recommend one mode or another for future studies; rather, we provide information to decision makers who must weigh a broader set of trade-offs and constraints.

The official charge reads:

> *A change in survey mode will have implications for the implementation of the ANES and, potentially, for the resulting survey estimates. The ANES has already taken important steps in providing evidence about the implications of a potential mode change, most recently in 2008 and 2012, with more to be learned from 2016. Others have conducted a variety of other studies that provide information along these lines. This committee is to ask what we have learned from what we (and from what others) have done. What do we already know? For example, in terms of potential differences in sample estimates, question wording, DK/NA options, and other design decisions? Where are we to go from here? Since in practice the last two election studies are a mixed mode design (and of course there are many other examples of mixed modes) how might the mixture of modes be leveraged effectively? What needs to be done to make a scientifically sound transition to some other mode or modes, if that is deemed desirable or necessary? What is the science that most needs to be done? How can we move forward while the science is under study?*

The committee includes D. Sunshine Hillygus (Duke University), chair, Nicholas Valentino (University of Michigan), Lynn Vavreck (University of California, Los Angeles), Matt Barretto (University of California, Los Angeles), and Geoff Layman (Notre Dame). The committee work was assisted by Brian Guay (Duke University), Kirill Zhirkov (University of Michigan), and Jack Zhou (Duke University).

Introduction

Since its inception, the ANES has been conducted using an area-probability sample and face-to-face interviews -- a design that is thought to produce the highest quality data but is increasingly expensive as response rates have declined in recent decades (Singer 2006, Groves 2006, National Research Council 2013). As summarized in *The Future of Survey Research*, a 2012 report to the National Science Foundation Advisory Committee for the Social, Behavioral and Economic Sciences Subcommittee on Advancing SBE Survey Research: "research to date suggests that face-to-face interviewing may be the method most likely to generate the highest response rates, the greatest trust and rapport between the researchers/interviewers and the respondents, the most cognitive effort from respondents in generating answers accurately, and the most honesty when providing reports regarding sensitive topics. But face-to-face interviewing is extremely expensive, and the costs of implementing such efforts well have been rising quickly" (Krosnick et al. 2012, 28).

Increasing costs, declining response rates, changes in funding environments, and technological advances, however, have led many large-scale survey projects in government, academia, and industry to consider alternative design approaches. Many of these new designs include self-administered questionnaires.

In writing this report, the committee used the following elements to evaluate these opportunities and challenges: committee discussions at Board meetings, an extensive literature review of the literature on mode, analysis of the 2012 ANES that had parallel self-completed and face-to-face modes of interview, replication of these comparisons where possible with the 2016 ANES data, and a comparison of mode effects from a randomized mode experiment in Las Vegas. It is important to note that in the 2012 and 2016 ANES the change in mode is also conflated by a change in sampling methodology. In 2012 the self-completed interviews were done as part of an existing online survey panel. In 2016, the self-administered surveys were given to a randomly selected set of people selected by a multiple-frame probability sampling method.

As will be discussed in the results section below, differences in the design of the online surveys in 2012 and 2016 provide some leverage for evaluating how different online survey design decisions might affect data quality. It is also worth noting that the ANES conducted a dual-mode design in 2000 in which a portion of the sample was completed by telephone, but results from experiments suggested that an all-telephone survey was not a preferred alternative (Bowers and Ensley 2003). Nonetheless, telephone interviewing could still be considered as a potential component of an explicitly multi-mode design.

To summarize, the key recommendations of the report are as follows:

1. We believe it is critical that any online implementation of the ANES be conducted using a freshly-recruited probability sample of the electorate (i.e., not sampled from a panel of frequent survey takers).

2. A mode change inevitably creates tension between comparability across modes due to distinct questionnaire construction, respondent demand, sampling variation

and the specific best practices for measurement error reduction associated with a given mode. That is, simply and directly implementing the in-person questionnaire for the online sample might maximize comparability, but it comes at the cost of data quality. We believe further research is needed to more fully understand these differences and ensure comparability and optimization for mode. For example, modularization of the online survey instrument into a series of shorter questionnaires is probably necessary to achieve high completion rates. In addition, volunteer response options (e.g., "Don't Know" for some questions) have to be reconsidered for online implementation. Design decisions also need to explicitly consider mobile device completion for online surveys.

3. Given differences in response biases across modes, researchers using the 2012 and 2016 samples should explicitly evaluate differences in their results across these samples, particularly when the construct being measured is associated with known differences between the two samples (e.g., political knowledge, evaluations of groups or candidates, and engagement).

Background

Considering potential concerns about the funding environment and the increasingly widespread use of self-completed surveys in published political science research, it is natural to evaluate alternative designs for the ANES. Within the discipline, online surveys are an increasingly prevalent source of data in published research. A small number of comparison studies conclude online surveys – with self-completed questionnaire's – offer acceptable representative samples of the general population (Ansolabehere and Schaffner 2014, Sanders et al. 2007, Rivers 2007, Berinsky et al. 2012). Most pointedly, Ansolabehere and Schaffner (2014) conclude that "a carefully executed opt-in Internet panel produces estimates that are as accurate as a telephone survey and… the two modes differ little in their estimates of other political indicators and their correlates."

A review of the broader survey methodology literature provides more nuance. This literature emphasizes the extent of variation in online survey design and quality, some of it arising from the sampling methodology associated with this particular mode of interview. An extensive body of work identifies potential concerns with measurement and coverage error associated with some online methods (Rogers et al. 2005, Chang and Kronsick 2009, Yeager et al. 2011, Atkeson et al. 2014, Pasek and Krosnick 2010, Malhotra and Krosnick, 2007; Yeager et al., 2009,Newman et al. 2002, Des Jarlais et al. 1999, Riley et al. 2001, Cooley et al. 2001, Metzger et al. 2000, Waruru, Nduati, and Tylleskar 2005, Ghanem et al. 2005; Taylor, Krane, and Thomas, 2005; Crete and Stephenson, 2008; Braunsberger, Wybenga, and Gates, 2007; Klein, Thomas, and Sutter, 2007; Thomas, Krane, Taylor, & Terhanian, 2008; Baker, Zahs, and Popa, 2004; Schillewaert and Meulemeester, 2005; Roster, Rogers, Albaum, and Klein, 2004;

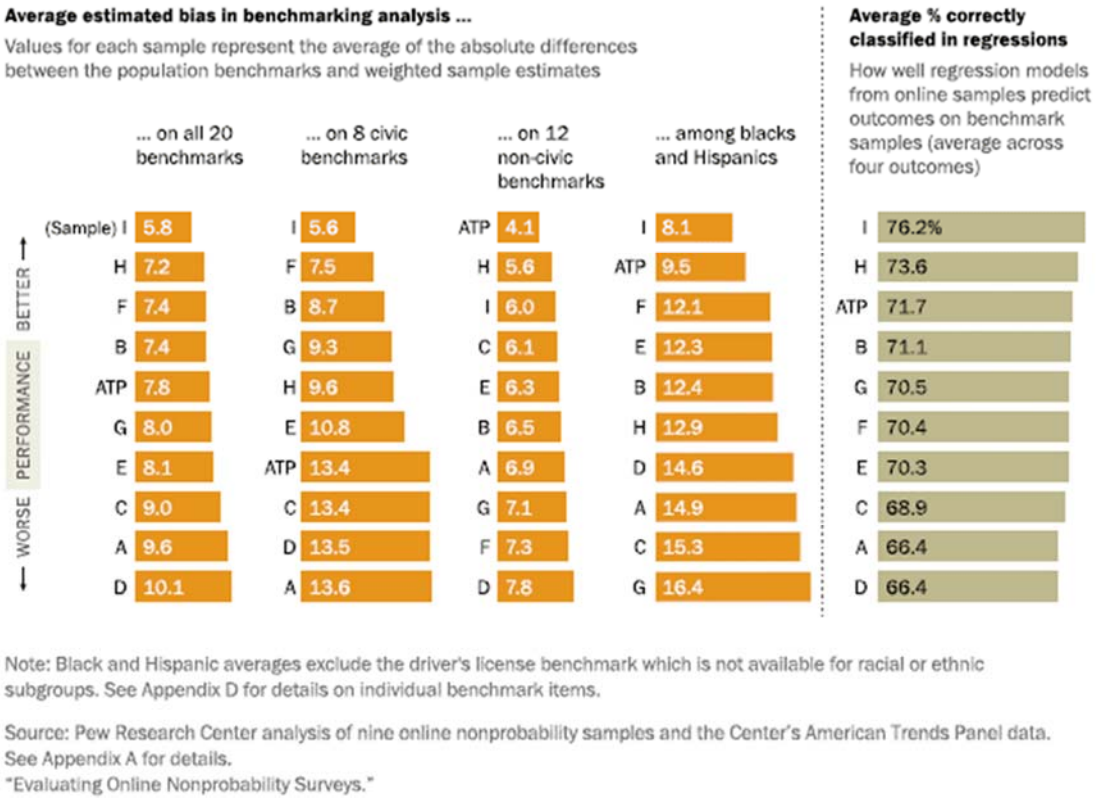
Loosveldt and Sonck, 2008; Miller, 2000; Burke, 2000; Niemi, Portney, and King, 2008; Schonlau, Zapert, Simon, Sanstad, Marvus, Adams, Spranca, Kan, Turner, and Berry, 2004; Berrens, Bohara, Jenkins-Smith, Silva, and Weimer, 2003; Sparrow, 2006, Cooke, Watkins, and Moy, 2007; Elmore-Yalch, Busby, and Britton, 2008; Niemi, Portney, and King, 2008). We briefly summarize some of the key issues that have been raised. For an extensive review, see Baker et al. (2010) and Callegaro et al. (2014).

First, there is significant variability in data quality across on-line survey vendors. Because there is no sample frame of email addresses, many online surveys rely on online panels, most often non-probability, opt-in panels. Without a probability sample and high response rate, models and assumptions are required to make inferences to a broader population (e.g., ignorability, exchangeability, post-stratification weighting). Unfortunately, we rarely have the information necessary to fully evaluate the effectiveness of these models for reducing bias and there is considerable variation in the models used by different online panels (Baker et al. 2010). Because of this, some take the rigid view that there "is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. Thus, claims of 'representativeness' should be avoided when using these sample sources" (Baker et al. 2010). Others conclude that an online non-probability sample's fitness-for-use depends on the specific application of interest (and on the judgment of the researcher and reviewers), but emphasize the necessity of transparency about the data collection and recommend that accuracy be explicitly evaluated to the extent possible (Baker et al. 2010; Vanette and Krosnick forthcoming).

A recent Pew Study illustrates this wide variability across online panels. Nine well-known non-probability panels (labeled with letters to protect confidentiality) and one probability-based panel (Pew's American Trends Panel (ATP), which relies on GFK Knowledge Networks) are evaluated.[1] The analysis found considerable variability in terms of quality and accuracy against known benchmarks (Kennedy et al. 2016).

[1] For more, see Doug River's blog post regarding the report: https://today.yougov.com/news/2016/05/13/pew-research-yougov/.

Figure 1: Variances in benchmarks across various online surveys



Source: Kennedy et al. (2016)

Even when a panel is initially recruited using probability methods, as with the ATP or the 2012 ANES, population estimates have been found to be biased relative to benchmarks. Several different sources of bias have been proposed. Those who agree to participate in a panel, taking frequent surveys in exchange for money or gifts, could differ in consequential ways from those who decline to participate. Those who drop-out of the panel over time could differ from those who stay in the panel (panel attrition). Or participation in frequent survey taking could change the attitudes or behaviors of panelists (panel conditioning). For example, Sturgis, Alum, and Brunton (2009) find that frequent survey takers tend to have stronger and more internally-consistent attitudes. These are patterns that we see among the 2012 ANES online respondents, who had previously taken an average of 324 surveys as part of the GfK panel.

Based on our literature review and the original analysis presented in the next section, the committee *recommends that the ANES prioritize a design that relies on a fresh, probability sample, regardless of interview mode*. Such a commitment may have cost implications that should be explicitly considered.[2]

[2] Although we are evaluating the mode change in this report with respect to the continuity of the time series, it is worth mentioning that the reliance on an in-person interview necessitates the area-probability sampling method the ANES has used for decades. This clustered sampling method increases the design effect of the survey relative to a

A second issue in considering a shift from in-person interviews to self-completed surveys is measurement error. The presence of an interviewer changes the dynamics of the data collection, with both positive and negative measurement implications. Considerable research has shown that self-administered computer-based interviews find higher levels of stigmatized behaviors and attitudes than face-to-face interviews (e.g., Kreuter, Presser, and Tourangeau, 2008). In other words, online surveys should reduce social desirability biases in responses. At the same time, respondents may be less likely to remain engaged and attentive in the absence of an interviewer. That is, a key disadvantage of online surveys is that they have been shown to have greater evidence of satisficing—respondents failing to expend the cognitive effort to thoughtfully answer survey questions. Respondents introduce measurement error if they do not take a survey seriously and simply speed through the questions to get to the end. For example, research has found both higher levels of "don't know" and non-differentiation (Chang and Krosnick; Heerwegh and Loosveldt 2008), and speeding (Chang and Krosnick; Heerwegh and Loosveldt 2008; Miller 2000; Burke, 2000) in self-completed interviews than face-to-face interviews. . Our analysis of the ANES finds response patterns consistent with some of these measurement biases, although the comparison is sometimes conflated with questionnaire and sample differences.

It is also worth noting that there are many other sources of measurement error associated with each mode that we don't fully consider in this report. For example, recency effects are more likely to occur for long lists of response options when read by an interviewer, primacy effects are more likely to occur in self-complete modes (Tourangeau, Rips, & Rasinski, 2000). The race and gender of the interviewer can bias results in an in-person interview; The visual display shapes how people respond in a self-complete survey (Dillman 2011). These and other factors can shape the answers people give and thus the comparability of estimates across modes.

There is an emerging literature on design best practices to help mitigate measurement error with each mode (e.g., Dillman & Christian 2005, Dillman 2011, Smyth et al. 2006). For example, online surveys often exclude an explicit "don't know" category, whereas surveys with an interviewer can allow the respondent to volunteer a "don't know" response (Duffy et al. 2005). Similarly, it is known that longer surveys increase satisficing behavior, so an online survey might be modularized to reduce satisficing related to fatigue (Galesic & Bosnjak 2009). However, while optimizing surveys for the mode of completion can reduce measurement error, it can also reduce comparability across modes. This tension has been studied most explicitly with

---

simple random sample, while also increasing the costs associated with travel for the interviewers. The ANES staff estimate that the cost-per-case in the 2016 face-to-face sample was approximately $3050.00 exclusive of staffing and overhead costs. The cost-per-case in the 2016 self-completed probability sample was approximately $400.00. ANES staff estimated the costs for the 2012 interviews at approximately $2,000 per in-person respondent and $230 per online respondent. Cost-per-case is of course dependent on the total sample size relative to fixed project costs, excluding staffing and overhead costs. Specific design decisions will of course alter these cost differences for future studies. We do not explicitly consider cost-quality trade-offs in this report.

respect to multimode designs, an increasingly common approach used to reduce costs and increase response rates (de Leeuw 2005). For example, surveys can begin with a less expensive mode of data collection (to reduce cost) and then switch to more expensive modes (to reduce non-response). Dillman and Christian (2005) seem to suggest that it is best to maximize comparability between modes in a mixed-mode design through use of a unimode design for the questionnaire that avoids design features not replicable in every mode used. In contrast, Groves and Heeringa (2006) suggest optimizing each questionnaire to reduce measurement error associated with a given mode, even if it reduces comparability across modes. Our conclusion is that *more research is needed to determine how to best balance the tension between comparability across modes (and years) and optimization of the questionnaire for mode.*

Original Data Analysis

To evaluate the potential challenges and benefits of including a self-completed mode of interview in conjunction with or as a replacement to the traditional in-person interviewing, we analyze 3 sets of data: the 2012 and 2016 ANES time series studies and a 2011 mode experiment. In 2012 and 2016, the ANES conducted parallel surveys with both online and in-person samples; because all samples were probability-based, both the online samples and the in-person samples have been touted as representative of adult, non-institutionalized citizens in the continental United States. To be sure, because mode was not randomized in these years, the comparison of these samples cannot cleanly distinguish mode effects from sample differences (e.g., differences in response rates or differential attractiveness of a given mode for a given respondent type), but the questionnaires were administered with (mostly) consistent question wording and response options. Moreover, the differences in the design of the two online samples—while holding constant the face to face method—offers some leverage to distinguishing sample from mode effects.[3]

In 2012, the online sample was drawn from Knowledge Network's KnowledgePanel, an online panel that uses probability-based random digit dialing (RDD) and address-based sampling techniques to obtain a national sample of American adults with and without internet access. The sample sizes for the two modes were 2,054 respondents for in-person and 3,860 respondents for online (total N = 5,914). In 2016, the online ANES survey employed a "fresh recruit" design drawn from multiple frames including the US Postal Service's computerized delivery sequence file (DSF), which includes residential addresses from each of the 50 states and the District of Columbia. The sample size for the 2016 in-person survey was 1,181 respondents and for the online survey was 3,090 respondents (total N = 4,271).

A small number of studies have examined specific differences between modes on the 2012 ANES. Most comprehensively, Homola, Jackson, and Gill (2016) develop a new statistical

[3] The 38% response rate of the 2012 face-to-face survey is the lowest in the history of the project. It seems plausible that nonresponse bias could have contributed to some of the observed differences.

estimate of the overall distributional difference across the two surveys in entirety and conclude that online sample was more likely to have extreme responses. In a more focused comparison, a series of studies finds that the online sample has warmer feeling thermometer ratings (Liu and Wang 2015) and higher levels of political knowledge (Liu and Wang 2014). Karp and Luhiste (2015) specifically focus on differences in the relationship between age and civic engagement across the samples, finding that the online sample overestimates civic engagement, producing a different explanation for what motivates people to vote.

Building on this initial work, we examine the 2012 and 2016 versions of the ANES and compare how the in-person survey mode in each year corresponds to its online counterpart, though the 2016 analysis is still quite preliminary due to the recent release of the data. We briefly summarize the findings in the text and provide more detailed tables in an appendix.

Demographic Comparison

We begin by examining differences between samples on demographics and important political variables describing respondents' political preferences, attitudes, and voting behavior. We present a comparison of demographic variables across samples in Table 1a for the 2012 survey and Table 1b for the 2016 version, reporting both the weighted and unweighted means for the in-person and online samples.[4] To briefly summarize, there are significant differences between the raw samples that are corrected once post-stratification weights are used. These differences are especially pronounced in the 2012 sample, though the 2016 sample does appear to oversample highly educated and wealthy individuals. Post-stratification weighting mostly corrects the sample average for the demographic variables on which the weights were constructed, but that is by design since the two samples are being matched to the same "truth benchmarks." However, research has shown that demographic variables are often insufficient to account for nonresponse and coverage biases (e.g., Tourangeau, Conrad, and Couper, 2013). It is also important to scrutinize joint distributions; although internet penetration rates now exceed 88% of the population,[5] there remains disparities in access and quality of access by race/ethnicity, age, and socio-economic status. This may be less of a concern in the 2016 approach, since the person does not have to have access to the Internet at home to complete the survey (for example, they are instructed to go to a library or to call to request a paper version of the survey).[6] Adaptive design is an increasingly popular approach to try to improve the representativeness of samples through the survey design (e.g., through use of differential incentives and follow-ups) (LeFlamme and Wagner Forthcoming).

Political Engagement

[4] For ease of presentation, we use means throughout the descriptive comparisons. See Homola, Jackson, and Gill (2016) for discussion of alternative metrics of centrality and dispersion.

[5] http://www.internetlivestats.com/internet-users/us/

[6] Further concern should be given to cultural variation in how the Internet is accessed in the home. For example, in many Latino households, access is via mobile device, which may not be compatible with a survey platform's online capabilities. Similarly, illiterate citizens may have a hard time completing even if they would like to comply.

In Tables 2a and 2b, we report several indicators of political engagement. While we observe no differences in how frequently respondents report paying attention to politics or following campaigns, online respondents were more likely to have heard or read about the 2012 presidential campaign online, in the newspaper, or on the radio than in-person respondents. Likewise, online respondents were 12 percentage points more likely to report knowing where to vote ($p < 0.001$), as well as 5 percentage points more likely to say that they are registered to vote ($p < 0.001$) and 12 percentage points more likely to report that they voted in a primary election ($p < 0.001$). These differences between survey modes are largely absent from the 2016 online sample, where the only significant difference among our selected variables was in respondents who reported being contacted by a political party during the campaign (28% in-person vs. 33% online, $p < 0.05$). The magnitude of the gap between in-person and online surveys on this item mirrors that of the 2012 ANES (39% in-person vs. 44% online, $p < 0.01$). The overall pattern of increased political attentiveness is consistent with previous research finding that individuals in online panels are more politically engaged than the general population—although it remains unclear if this reflects sample selection among those initially recruited, panel attrition among those who are uninterested, or panel conditioning from being asked so many political surveys. The 2016 sample of freshly recruited respondents does not show the same degree of asymmetry in interest or engagement as the 2012 panel-recruited sample.

Political Knowledge

Political knowledge is another measure in which there has previously been found large differences between online samples and samples using phone or in-person surveys (Fricker et al. 2005, Lui & Wang 2014). In both the 2012 and 2016 ANES comparisons we find significant differences between in-person and online samples. In particular, on less publicly salient items such as those asking about Senate term length and correctly identifying individuals' political positions, online respondents appeared considerably better informed than those in the in-person samples. As shown in Tables 3a and 3b, the percentage of correct identifications for respondents in the online samples show that they massively outperformed their in-person counterparts, at times more than doubling the in-person sample's performance. Given the consistency of this finding across 2012 and 2016, even though the latter sample does not consist of experienced survey respondents who are more politically engaged, it raises a question of whether online respondents are looking up answers to political knowledge questions. There is debate within the literature about the extent of cheating on political knowledge questions in online surveys (the alternate possibility being that because of the presence of an interviewer, respondents in in-person interviews are more likely to say "I don't know" to avoid embarrassing instances of guessing incorrectly even if they could make an informed guess) (Berinsky, Huber, and Lenz 2012; Ansolabehere and Schaffner 2014; Clifford and Jerit 2014; Gooch and Vavreck 2015), with considerable variability found across different types of online samples (Clifford and Jerit 2016). This again seems like an important area for further scrutiny and examination (e.g.,

examining completion time). There may also be design approaches that could help to alleviate such biases that are worth considering in future designs (Clifford and Jerit 2016).

Political Attitudes

We next examined the political attitudes held by respondents in the in-person and online samples of the two ANES studies. Among the most important political attitudes are party identification and ideology (table 4a and 4b), but the comparisons here also highlight the tension of optimizing for mode. Comparisons across the samples on these items are complicated by subtle changes in question wording for online implementation. For party identification, online respondents in 2012 were provided with the option of leaning towards neither party if independent was selected, while in-person respondents had to volunteer this option. Still, more online respondents selected "Neither Party" (10%) than in-person respondents who volunteered "neither party" or "Don't Know" (16%). Thus, there are significantly more Independents in the 2012 online sample (16%) than the in-person sample (10%). For comparability, we can instead use responses from the first party identification question only, when respondents are asked to identify as Democrat, Independent, or Republican. This still produces more independents in the in-person sample than the online sample (36% vs. 31%, $p < .01$), as well as fewer Republicans (24% in the in-person sample vs. 29% in the online sample, $p < .01$) but roughly the same number of Democrats.

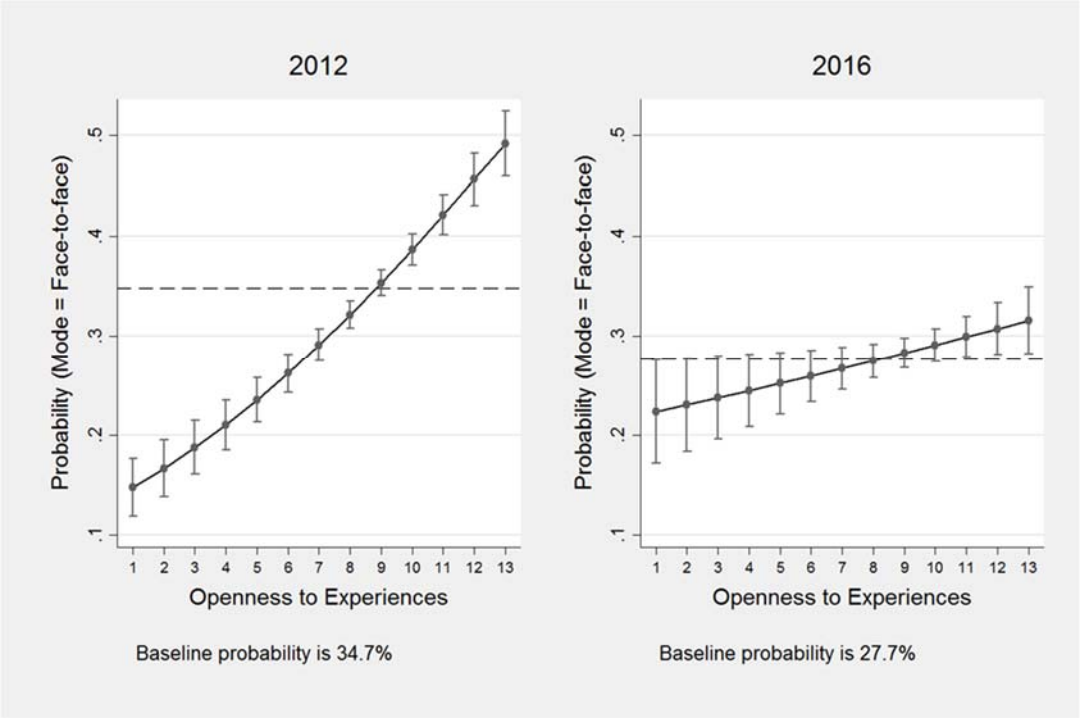
The ideology questions in 2012 also had small but consequential wording differences across the two samples. All respondents in both samples were asked to place themselves on a 7-point scale from "extremely liberal" to "extremely conservative." However, in-person respondents had two additional response categories: "Haven't thought much about this" and a volunteered response of "don't know" that could be recorded by the interviewer. While less than 1% of in-person respondents volunteered "don't know," 27% selected "haven't thought much about this." When these additional response options are excluded from the analysis, the online sample appears significantly more moderate (36% online vs. 31% in-person, $p < .01$). When these additional responses are coded as moderate, the opposite trend emerges; the online sample is less moderate than the in-person sample (36% online vs. 47% in-person, $p < .001$).

In the 2016 ANES, these differences in question wording between online and in-person surveys persist. For party identification, in-person respondents were given not only the option of "don't know" as in 2012, but additionally a "no preference" choice option. For political ideology, in-person respondents were offered "don't know" and "haven't thought much about this" options, with interviewers instructed not to probe this second choice of non-answer. The 2016 online sample reports being significantly more Independent (22% online vs. 18% in-person, $p < 0.05$) and less Republican (38% online vs. 43% in-person, $p < 0.05$) than the in-person sample. However, despite the discrepancies in question wording for ideology, the online sample does not significantly differ from the in-person sample on this dimension.

In addition to a comparison of ideology and party identification, we examined several other political attitudes. Although not everything is shown here, we have scrutinized both individual items and various scales. Differences in issue scales are summarized in Tables 5a and 5b. In the 2012 ANES, we find significant differences between the two modes on many of these attitudinal items. In terms of the observed patterns, some seem consistent with response biases across modes (e.g., government spending items), while others seem to reflect sample composition or may simply reflect survey noise. Most notably, the online sample in 2012 more often offers conservative policy attitudes than the face-to-face sample. The differences are at times small and insignificant, but there are sizable differences in the conservative direction for online respondents on an omnibus scale of issue placements as well as specific policies like immigration, the environment, and a summary of spending on social welfare and other liberal government initiatives. In the 2016 ANES, the in-person and online samples are much more similar, although online respondents still favored less government spending than in-person respondents, but the differences were substantially smaller.

On further analysis, we suspect these differences reflect sample composition; that is, differences in the type of person who is willing to answer hundreds of surveys alone online compared to the type of person who is willing to allow a stranger into their home to complete a 90 minute face-to-face survey. In 2012, we find large differences in openness to experience and extroversion across the two samples. Openness is correlated with liberal policy views, and the face to face sample was substantially higher in openness. As shown in the left-hand panel of Figure 2, participation by mode was strongly correlated with openness to experience in 2012. In 2016, however, that effect was greatly diminished. This would lead us also to predict that ideological differences across mode should be smaller in 2016, and indeed they were.

Figure 2. Mode of Completion (predicted) by Openness to Experience



Note: Plots present predicted marginal effects of Openness to Experiences on the mode of completion from the logistic regressions in 2012 and 2016 ANES. Modelled effects controlled for other Big Five personality traits. For computation of marginal effects, all traits other than Openness to Experiences were kept at their respective means. Dashed line is the observed proportion of face-to-face respondents in the sample.

The greater similarity in political attitudes between the online and in-person samples of the 2016 ANES again point to the higher data quality attainable by the random recruitment sampling method than by using existing online panels.

Candidate Assessments

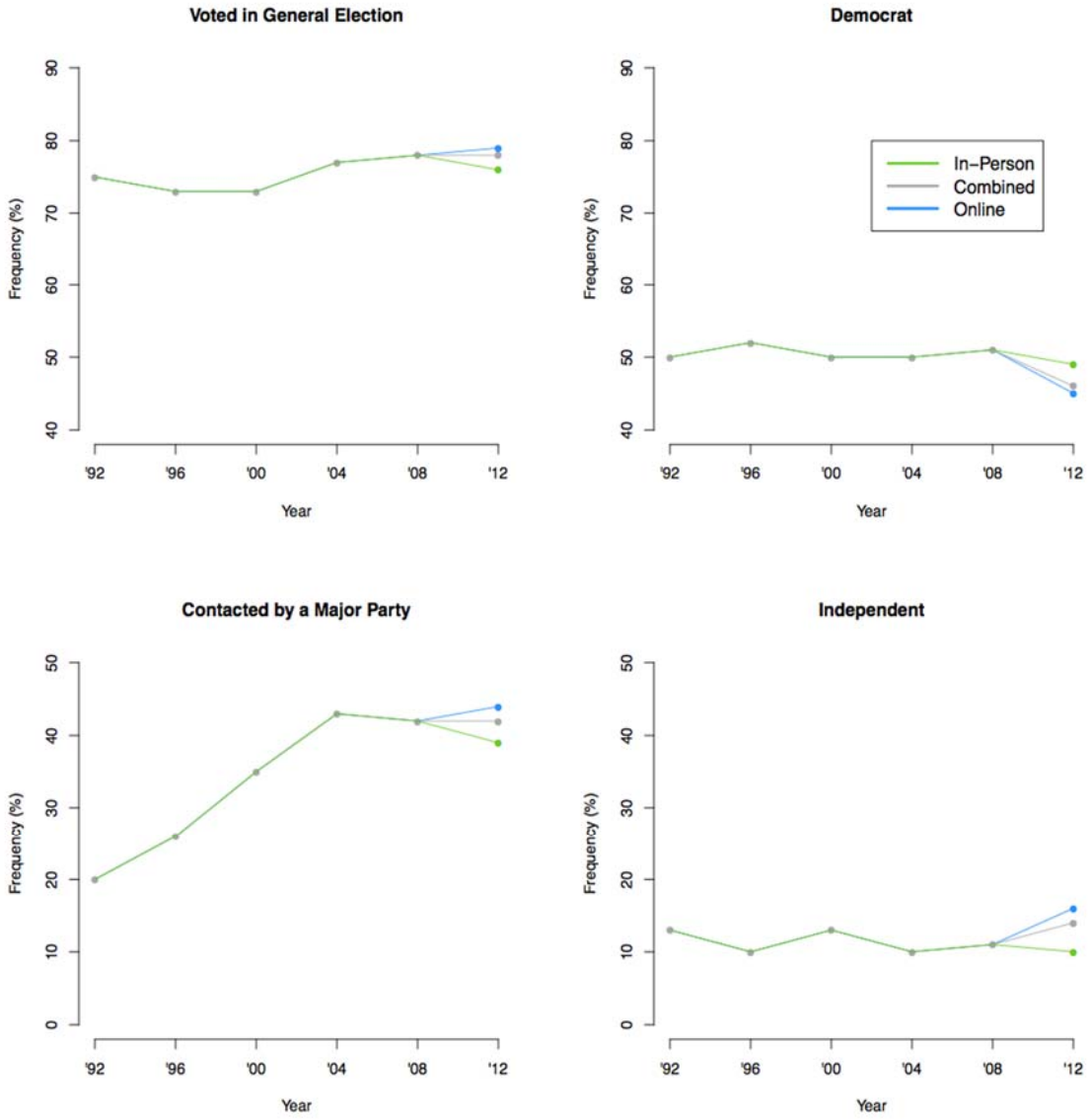
In both the 2012 and 2016 versions of the ANES, respondents were asked to assess the presidential candidates of the two major parties across a range of perceived character traits, including leadership, empathy, and honesty. In both cases, online respondents reported less favorable impressions **of both candidates** than respondents in the in-person modes (Tables 6a and 6b). For instance, in the 2012 ANES, respondents reported significantly more positive ratings on Barack Obama's leadership, knowledge, empathy, and honesty traits as well as Mitt Romney's knowledge. These differences between modes were less pronounced in the 2016 ANES, but the same pattern of higher in-person candidate trait assessments can be observed in Hillary Clinton's perceived knowledge (3.55 in-person vs. 3.37 online, $p < 0.01$ on a 5 point scale) and penchant to speak her mind (3.04 in-person vs. 2.86 online, $p < 0.01$), as well as Donald Trump's perceived knowledge (2.44 in-person vs. 2.28 online, $p < 0.05$) and penchant to speak his mind (4.28 in-person vs. 4.14 online, $p < 0.05$). In light of the other differences observed across samples, we suspect this pattern could reflect respondents being unwilling to

appear too negative when answering questions with a live interviewer, although it could also reflect the greater percentage of independents/moderates in the online sample.

Mode Differences: Trends in Public Opinion

To illustrate the implication of the mode differences, we provide four examples in which interpretation of public opinion trends over time are impacted by the sample analyzed. For instance, in 2012, self-reported voter turnout decreases substantially from the long-term trend when using the in-person sample only, increases when using the online sample, and remains stable compared with recent elections when combining both samples.

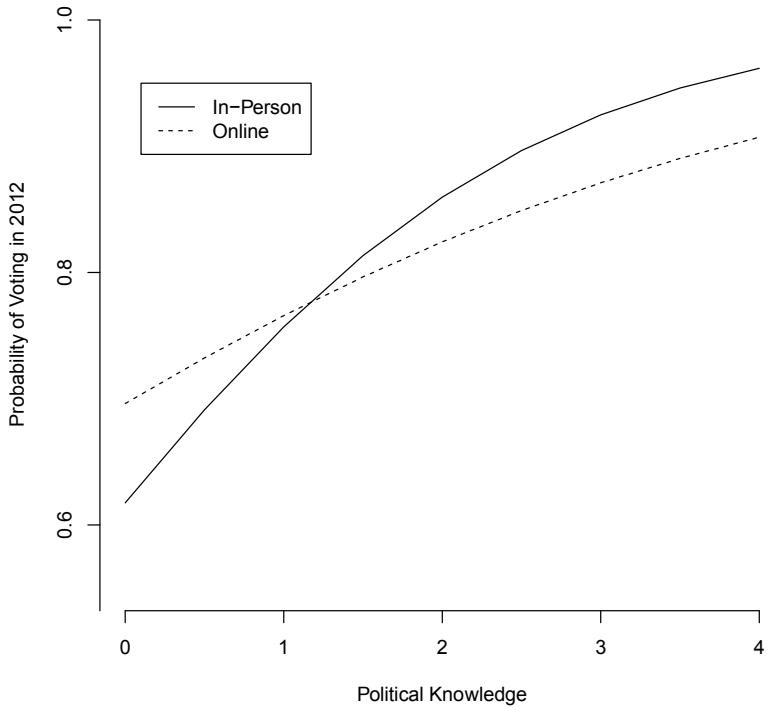
Figure 3: Political Trends across Elections



Note: Voting behavior across the past 6 presidential elections by mode. All means are weighted. Sample differences for these variables are presented in Tables 3 and 4 above.

We have also identified cases in which the samples result in different results within a multivariate setting. The relationship between political knowledge and turnout differs depending on the sample. Importantly, simply including a control for mode was insignificant and was insufficient for accounting for the difference identified with an interaction in the mode. [7]

Figure 4: Probability of Voting Across Political Knowledge (2012)



Note: The results from the full model are reported in the appendix. Knowledge is measured as an additive scale representing the number of correct office recognition items (see Table 3a). Online is an indicator variable representing mode, where 1 indicates online and 0 indicates in-person.

[7] The four office recognition items on the ANES were combined into a single variable, which represents the total number of correct answers to the office recognition questions. Rather than arbitrarily holding values of covariates at their central tendencies, we ran simulations to calculate the predicted probabilities, using the actual values of each respondent to produce coefficient estimates and their associated errors. We simulated 1,000 coefficient estimates from the model, calculated a predicted probability of voting for each respondent using their actual values of each predictor, and averaged over across these probabilities to obtain the mean predicted probability for each respondent. Finally, we took the mean of these probabilities to obtain the plotted probabilities and associated confidence bounds.

2011 Mode Experiment

Overall, the ANES comparison of the in-person and online versions of the 2012 ANES reveals substantive and significant differences in political attitudes and behavior of the two samples. However, because the 2012 ANES simultaneously altered both the interview mode and sampling design we are unable to determine from which of these changes the differences originate. The additional analysis of the 2016 mode comparison, in which the online same replied on a freshly-recruited online probability sample used offers suggestive evidence. A comparison of the online and in-person samples on the 2016 ANES revealed fewer differences in political preferences and attitudes than the 2012 ANES, with political knowledge being the only exception. This comparison suggests that many of the differences observed in 2012 are likely a reflection of sampling differences rather than mode effects.

This conclusion is supported by a cursory analysis of a randomized mode experiment conducted by Lynn Vavreck in Summer 2011 at Television City, the CBS research facility inside the MGM Grand Hotel in Las Vegas, Nevada. Respondents were randomly assigned to complete a political survey in the presence of a live interviewer or on a desktop computer.[8] This randomization allows isolation of the mode effect without the confounder of sample differences. While the experiment only measured a subset of the many policy preferences and political attitudes featured on the ANES, our initial analysis found fewer differences in political attitudes and preferences on this survey than on the 2012 ANES. This lends support to our first recommendation that any alternative mode of data collection on the ANES features a freshly recruited sample as opposed to a sample drawn from an existing online panel. When looking at differences due to mode alone, the data looked more like the 2016 comparison (with a newly recruited probability sample) than the 2012 differences (that relied on an existing on-line panel).

The mode experiment also lends support to our remaining recommendations. First, the mode experiment finds evidence of higher levels of socially desirable answers in the in-person condition (Table 8b). Though these differences are only statistically significant in 2 of the 6 items, those two items (Mormons and Muslims) were the most sensitive.

Finally, the mode experiment highlights the need to better understand political knowledge differences in online and in-person samples. The analysis found that respondents in the self-complete condition provided more correct responses on all 10 general knowledge and political knowledge items, though these differences were only significant for about a third of questions (see Table 8a). This is a much smaller gap than what was observed in the ANES comparisons. It may be the case that subjects were less inclined to cheat in the Las Vegas experiment than when completing the survey at home. Further research is needed to understand the causes of the knowledge discrepancy between modes with an eye towards identifying strategies to reduce biases.[9]

---

[8] Respondents were randomized 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). The experiment ran from July 26th to August 8 6th, 2011, and respondents were compensated with a $5 gift card.

[9] Although the knowledge gap here is far smaller than observed in the 2012 and 2016 ANES, its endurance despite the decreased probability of looking up correct answers is puzzling. One potential explanation is that the greater number of incorrect responses in the in-person mode is driven by a greater tendency to respond "don't know". Indeed, on the question about who holds the office of the Vice President, 63% of in-person respondents volunteered

Satisficing Across Modes

Previous research has suggested that self-completed samples have higher levels of satisficing than in-person samples, but the previous analysis does not evaluate if that is the case within the ANES data. Satisficing cannot be directly observed, but typical indicators include fast response times, use of "don't know" responses, item nonresponse, random response selection, open-ended gibberish, and straight-lining responses (Baker et al., 2010). Unfortunately, the 2012 and 2016 samples were not designed in a way that allow an ideal evaluation of the extent to which data quality might be affected by a move to self-completion because mode was not randomized and levels of satisficing can be shaped by mode, survey length, incentives, interviewer experience, etc. The 2012 mode comparison is especially muddied by the use of an online panel of frequent survey takers. The 2011 experiment had the necessary randomization—and analysis of item nonresponse across modes using the data offer encouraging results (Vavreck and Gooch 2016), but the questionnaire was not the full (lengthy) ANES instrument.

The 2016 ANES study offers a more relevant comparison, although some of the variables needed to calculate the informative metrics have not yet been released (e.g., questionnaire timing, open-ended responses). Some questionnaire differences also limit the items available for comparison; for example, for many items in-person respondents were allowed to volunteer a "don't know" response whereas online respondents did not have that choice. For the handful of questions in which "haven't thought much about this" was an explicit option for both samples, we do tend to find a higher percentage of online respondents choosing it. These comparisons are show in Table 9a. Overall, item nonresponse was also higher in the online sample than the in-person sample, but the overall levels were very low. The online sample did make skipping a question difficult—respondents were allowed to skip a question only after a follow-up prompt. More pronounced are differences in willing to offer a like/dislike of candidates and parties. Across the entire battery, online respondents were significantly less likely to say they had a like or dislike compared to in-person respondents. For example, 49% of in-person respondents said they liked something about the Republican candidate compared to 39% of online respondents. According to ANES staff calculations, there was a similar gap in item nonresponse to the most important problem questions; whereas 10% of in-person respondents failed to respond, the rate was 17% for online respondents.

Conclusion

With the goal of continuing the time series data collection, our analyses of two recent ANES studies (2012 and 2016) with mode differences and an NSF-funded randomized mode experiment (2011) lead us to conclude one thing with a good deal of certainty: self-completed surveys (however implemented) should be done using a sample drawn from probability methods specifically for the purposes of the project. Our other conclusions, that more work is needed to assess how best to optimize instruments across mode and that users of the data should analyze the existing data from different samples separately before making inferences are no less founded in our review of materials, but are of less direct consequence to the future of this project.

---

a "don't know" response, compared to only 27% of respondents in the self-complete condition. A similar pattern of higher "don't know" responses among the in-person sample is observed across the three candidate ideological placement questions.

Works Cited

Ansolabehere, Stephen, and Brian F. Schaffner. 2014. "Does Survey Mode Still Matter? Findings from a 2010 Multi-Mode Comparison." *Political Analysis* 22(3): 285–303.

Atkeson, Lonna Rae, Alex N. Adams, and R. Michael Alvarez. 2014. "Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys." *Political Analysis* 22(3): 304–20.

Baker, R., Brick M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A., Krista, J.A., Gile, J., & Tourangeau, R. 2013. "Summary Report of the AAPOR Task Force on Non-Probability Sampling." *Journal of Survey Statistics and Methodology* 1(2): 90–143.

Baker, Reg et al. 2010. "Research Synthesis: AAPOR Report on Online Panels." *Public Opinion Quarterly* 74(4): 1–71.

Berinsky, A., G. Huber, and G. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon. Com's Mechanical Turk." *Political Analysis* 20: 351–68.

Bowers, Jake, and Michael J. Ensley. 2003. *Issues in Analyzing Data from the Dual-Mode 2000 American National Election Study*.

Callegaro, M., Baker, R.P., Bethlehem, J., Göritz, A.S., Krosnick, J.A. and Lavrakas, P.J. eds., 2014. *Online panel research: A data quality perspective*. John Wiley & Sons

Chang, Linchiat, and Jon A. Krosnick. 2009. "National Surveys via RDD Telephone Interviewing versus the Internet: Comparing Sample Representativeness and Response Quality." *Public Opinion Quarterly* 73(4): 641–78.

Cooley, P. C. et al. 2001. "Using Touch Screen Audio-CASI to Obtain Data on Sensitive Topics." *Computers in Human Behavior* 17(3): 285–93.

Des Jarlais, Don C. et al. 1999. "Audio-Computer Interviewing to Measure Risk Behaviour for HIV among Injecting Drug Users: A Quasi-Randomised Trial." *Lancet* 353(9165): 1657–61.

Dillman, Don A., and Leah Melani Christian. 2005. "Survey Mode as a Source of Instability in Responses across Surveys." *Field Methods* 17(1): 30–52.

Frankel, Laura Lazarus, and D. Sunshine Hillygus. 2014. "Looking beyond Demographics: Panel Attrition in the ANES and GSS." *Political Analysis* 22(3): 336–53.

Ghanem, K G et al. 2005. "Audio Computer Assisted Self Interview and Face to Face Interview Modes in Assessing Response Bias among STD Clinic Patients." *Sexually Transmitted Infections* 81(5): 421–25.

Groves, Robert M. 2006. “Nonresponse Rates and Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70(5): 646–75.

Groves, Robert M., and Steven G. Heeringa. 2006. “Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs.” *Journal of the Royal Statistical Society. Series A: Statistics in Society* 169(3): 439–57.

Hillygus, D. Sunshine, Natalie Jackson, and McKenzie Young. 2014. “Professional Respondents in Nonprobability Online Panels.” In *Online Panel Research: A Data Quality Perspective*, eds. Mario Callegaro et al. Chichester, UK: John Wiley & Sons, Ltd, 219–37.

Homola, Jonathan, Natalie Jackson, and Jeff Gill. 2016. “A Measure of Survey Mode Differences.” *Electoral Studies* 44: 255–74.

Karp, Jeffrey A., and Maarja Lühiste. 2016. “Explaining Political Engagement with Online Panels.” *Public Opinion Quarterly* 80(3): 666–93.

Kennedy, Courtney et al. 2016. *Evaluating Online Nonprobability Surveys*. Pew Research Center. http://www.pewresearch.org/2016/05/02/evaluating-online-nonprobability-surveys/

Leeuw, Edith D De. 2005. “To Mix or Not to Mix Data Collection Modes in Surveys.” *Journal of Official Statistics* 21(2): 233–55.

Laflamme, F. and J. Wagner. Forthcoming. Responsive and Adaptive Collection Designs. In Sage Handbook of Survey Methodology, eds. Christof Wolf, Dominique Joye, Tom W. Smith and Yang-Chih Fu.

Liu, Mingnan, and Yichen Wang. 2014. “Data Collection Mode Effects On Political Knowledge.” *Survey Methods: Insights from the Field*. http://surveyinsights.org/?p=5317.

Liu, Mingnan, and Yichen Wang. 2015. “Data Collection Mode Effect on Feeling Thermometer Questions: A Comparison of Face-to-Face and Web Surveys.” *Computers in Human Behavior* 48: 212–18.

Metzger, David S. et al. 2000. “Randomized Controlled Trial of Audio Computer-Assisted Self-Interviewing: Utility and Acceptability in Longitudinal Studies.” *American Journal of Epidemiology* 152(2): 99–106.

Montgomery, Jacob M., and Josh Cutler. 2013. “Computerized Adaptive Testing for Public Opinion Surveys.” *Political Analysis* 21(2): 172–92.

National Research Council. 2013. *Nonresponse in Social Science Surveys*. Washington, D.C.: National Academies Press.

Newman, Jessica Clark et al. 2002. “The Differential Effects of Face-to-Face and Computer Interview Modes.” *American Journal of Public Health* 92(2): 294–97.

Pasek, Josh, and Jon A. Krosnick. 2010. “Optimizing Survey Questionnaire Design in Political Science: Insights from Psychology.” In *The Oxford Handbook of American Elections and Political Behavior*, ed. Jan E. Leighley. New York.

Riley, E. D. et al. 2001. “Use of Audio Computer-Assisted Self-Interviews to Assess Tuberculosis-Related Risk Behaviors.” *American Journal of Respiratory and Critical Care Medicine* 164(1): 82–85.

Rivers, Douglas. 2007. *Sampling for Web Surveys.* Paper Presented at Joint Statistical Meetings. Salt Lake City, Utah.

Rogers, S M et al. 2005. “Audio Computer Assisted Interviewing to Measure HIV Risk Behaviours in a Clinic Population.” *Sexually transmitted infections* 81(6): 501–7. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1745071&tool=pmcentrez&rendertype=abstract.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, and Paul Whiteley. 2007. “Does Mode Matter for Modeling Political Choice? Evidence from the 2005 British Election Study.” *Political Analysis* 15(3): 257–85.

Schaeffer, Nora Cate, and Jennifer Dykema. 2011. “Questions for Surveys: Current Trends and Future Directions.” *Public Opinion Quarterly* 75(5): 909–61.

Singer, Eleanor. 2006. “Nonresponse Bias in Household Surveys.” *Public Opinion Quarterly* 70(5): 637–45.

Sturgis, Patrick, Nick Allum, and Ian Brunton-Smith. 2009. “Attitudes Over Time: The Psychology of Panel Conditioning.” In *Methodology of Longitudinal Surveys*, ed. Peter Lynn. Chichester, UK: John Wiley & Sons, Ltd, 113–26.

Waruru, Anthony K et al. 2005. “Audio Computer-Assisted Self-Interviewing (ACASI) May Avert Socially Desirable Responses about Infant Feeding in the Context of HIV.” *BMC Medical Informatics and Decision Making* 5(1): 24.

Yeager, David S. et al. 2011. “Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples.” *Public Opinion Quarterly* 75(4): 709–47.

Appendix

Table 1a: Demographic Comparison of Samples (2012)

	Unweighted Samples			Weighted Samples		
	In-Person	**Online**	**Difference**	**In-Person**	**Online**	**Difference**
Age	45.23 (.38)	51.62 (.26)	-6.40***	47.25 (.57)	47.44 (.38)	-.18
White	44.89 (1.10)	67.11 (.76)	-22.22***	70.89 (1.22)	70.94 (1.00)	-.05
Black	24.99 (.96)	13.15 (.55)	11.84***	11.93 (.81)	11.92 (.74)	.01
Hispanic	23.08 (.93)	13.88 (.56)	9.20***	10.88 (.73)	11.24 (.69)	-.36
Other Race	7.04 (.57)	5.86 (.38)	1.18	6.30	5.90	.41
College Degree	23.95 (.95)	35.02 (.77)	-11.08***	29.28 (1.41)	29.49 (.92)	-.20
Female	56.86 (1.10)	49.25 (.80)	7.62***	52.01 (1.53)	52.10 (1.06)	-0.10
Owns Home	54.31 (1.10)	72.74 (.72)	-18.44***	70.73 (1.25)	72.13 (.98)	-1.40
Income: 20k-49k	33.97 (1.07)	27.06 (.72)	6.90***	28.27 (1.33)	29.07 (1.00)	-0.80
Income: 50k-99k	24.41 (.97)	29.98 (.75)	-5.57***	29.95 (1.5)	31.94 (1.02)	-1.99
Income: 100k +	10.69 (.70)	20.80 (.66)	-10.11***	19.99 (1.42)	18.57 (.79)	1.41
Married	40.20 (1.08)	54.93 (.80)	-14.73***	53.25 (1.51)	53.25 (1.07)	0.00
Never Attend Worship	33.83 (1.05)	43.89 (.80)	-10.06***	37.24 (1.48)	45.96 (1.07)	-8.72***
No Religion	18.06 (.85)	22.93 (.68)	-4.87***	20.84 (1.24)	25.29 (.95)	-3.45**

Table 1b: Demographic Comparison of Samples (2016)

	Unweighted Samples			Weighted Samples		
	In-Person	**Online**	**Difference**	**In-Person**	**Online**	**Difference**
Age	49.8 (0.53)	49.5 (0.32)	0.4	47.66 (0.82)	47.18 (0.42)	0.5
White	78.09 (1.18)	80.19 (0.72)	1.11	78.71 (1.91)	77.73 (0.95)	0.98
Black	11.3 (0.92)	10.8 (0.56)	0.6	12.8 (1.41)	12.8 (0.87)	0.04
Hispanic	14.31 (1.02)	9.09 (0.52)	5.22	11.93 (1.4)	11.78 (0.67)	0.15
Other Race	5.42 (0.66)	5.28 (0.4)	.14	5.01 (0.93)	6.42 (0.61)	-1.42
College Degree	35.6 (1.39)	39.3 (0.88)	-3.8*	32.4 (2.77)	30.2 (0.86)	2.3
Female	51.7 (1.45)	52.5 (0.9)	-0.8	51.3 (1.86)	51.4 (1)	-0.1
Owns Home	59.4 (1.43)	62.8 (0.87)	-3.4*	60.2 (2.11)	59.8 (1)	0.4
Income: 20k-49k	27.9 (1.31)	25.1 (0.78)	2.7	26.1 (1.68)	24.6 (0.92)	1.4
Income: 50k-99k	29.1 (1.32)	30 (0.82)	.9	30.3 (1.47)	29.7 (1.1)	0.6
Income: 100k +	19.9 (1.16)	24.4 (0.77)	-4.5**	22.3 (1.83)	25 (0.88)	-2.8
Married	47.1 (1.45)	51.4 (0.9)	-4.3*	52.3 (1.86)	52.1 (1.02)	0.2
Never Attend Worship	36.36 (1.40)	43.00 (.89)	-6.64***	36.10 (1.74)	43.48 (1.07)	-7.38***
No Religion	16.04 (1.07)	22.49 (.75)	-6.45***	15.50 (1.26)	23.04 (.92)	-7.54***

Note: Means in the right panel are weighted using the full sample weight provided by the ANES. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 2a: Political Engagement (2012)

	In-Person	Online	Difference	Benchmark
Registered to Vote	84.75 (1.05)	89.42 (.71)	-4.67***	65.1 (CPS)
Voted in Primary	26.04 (1.37)	38.16 (1.01)	-12.12***	17.3 (FEC)
Voted in General Election	75.63 (1.31)	79.05 (.97)	-3.42*	54.87 (FEC) 56.6 (CPS)
Voted for Obama	52.48 (1.97)	51.00 (1.28)	1.48	51.06 (FEC)
Voted for Romney	45.87 (1.98)	45.02 (1.27)	.85	47.20 (FEC)
Contacted by Party	38.90 (1.56)	43.98 (1.08)	-5.08**	
Pays attention to politics (1-5)	3.34 (.04)	3.30 (.02)	.04	
Follows campaigns (1-3)	2.22 (.02)	2.26 (.02)	-.04	
Knows where to vote	69.04 (1.40)	81.38 (.92)	-12.34***	
Heard/read about campaign on:				
…TV	80.98 (1.25)	73.11 (1.02)	7.87***	
…Radio	33.38 (1.48)	38.05 (1.07)	-4.77**	
…Newspaper	45.64 (1.59)	50.65 (1.10)	-5.00**	
…Online	51.44 (1.58)	57.08 (1.10)	-5.63**	

Note: Means are weighted using the full sample weight provided by the ANES and are presented as percentages unless a scale is included in parentheses next to the variable name. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * p < .05, ** p < .01, *** p < .001. Benchmarks for voter turnout and vote choice from the Federal Election Commission[10], 2012 CPS Voting and Registration study[11], and the Bipartisan Policy Center.[12] Attention to politics is measured on a 5-point scale from "Never" (1) to "Always" (5), Follows campaigns is measured on a 3-point scale: "Not much interested" (1), "Somewhat interested" (2), and "Very much interested" (3).

[10] http://www.fec.gov/pubrec/fe2012/2012pres.pdf
[11] http://www.census.gov/data/tables/2012/demo/voting-and-registration/p20-568.html
[12] http://bipartisanpolicy.org/library/national-primary-turnout-hits-new-record-low/

Table 2b: Political Engagement (2016)

	In-Person	Online	Difference
Registered to Vote	85.4 (1.2)	83.6 (1)	1.7
Voted in Primary	39.8 (2.07)	43.6 (1.07)	-3.8
Voted in General Election	72.7 (1.97)	72 (1.08)	0.7
Voted for Clinton	49 (3.2)	48.8 (1.31)	0.3
Voted for Trump	42.1 (2.92)	44.2 (1.28)	-2.1
Contacted by Party	28.3 (1.61)	33.2 (1.29)	-4.9*
Pays attention to politics (1-5)	3.43 (0.04)	3.42 (0.02)	0.01
Follows campaigns (1-3)	2.35 (0.03)	2.35 (0.01)	0.01
Heard/read about campaign on:			
…TV news	68.11 (5.69)	73.02 (2.26)	-4.91
…TV talk shows	49.72 (6.23)	47.59 (2.27)	2.13
…Radio	41.19 (6.3)	40.66 (2.32)	0.52
…Newspaper	29.88 (6.55)	37.16 (2.34)	-7.29
…Online	43.29 (6.07)	49.41 (2.23)	-6.12

Note: Means are weighted using the full sample weight provided by the ANES and are presented as percentages unless a scale is included in parentheses next to the variable name. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Attention to politics is measured on a 5-point scale from "Never" (1) to "Always" (5), Follows campaigns is measured on a 3-point scale: "Not much interested" (1), "Somewhat interested" (2), and "Very much interested" (3).

Table 3a: Political Knowledge (2012)

	In-Person	Online	Difference
Majority party in House	51.14 (1.58)	68.00 (1.06)	-16.85***
Majority party in Senate	42.55 (1.58)	62.08 (1.09)	-19.53***
President can be elected twice	86.36 (1.00)	91.32 (.64)	-4.96***
Gov't spends least: Foreign Aid	31.64 (1.44)	32.47 (.98)	.83
Can Describe Medicare	70.51 (1.34)	82.19 (.86)	-11.68***
Senate Term Length	24.72 (1.37)	38.04 (1.02)	-13.32***
Correctly Identified			
…Speaker of the House	31.26 *(1.51)*	45.43 *(1.10)*	-14.17***
…Vice President	84.07 *(1.05)*	87.94 *(.77)*	-3.86***
... Chief Justice	14.60 *(.98)*	32.11 *(.94)*	-17.51***
… Prime Minister (UK)	10.26 *(1.01)*	23.50 *(.92)*	-13.25***

Note: Means are weighted using the full sample weight provided by the ANES and are presented as the percent of respondents who responded correctly to each political knowledge item. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3b: Political Knowledge (2016)

	In-Person	Online	Difference
Majority party in House	67.08 (1.48)	69.08 (0.96)	-2.01
Majority party in Senate	60.58 (1.5)	64.94 (1.04)	-4.36*
Gov't spends least: Foreign Aid	26 (1.58)	25.71 (0.89)	.3
Senate Term Length	22.76 (1.41)	44.52 (0.92)	-21.76 ***
Correctly Identified			
…Speaker of the House	43.55 (2.43)	57.81 (1.22)	-14.26***
…Vice President	81.92 (1.78)	88.35 (0.88)	-6.43**
... Chief Justice	12.93 (1.24)	40.7 (1.03)	-27.77***
… Chancellor of Germany	25.88 (1.95)	48.28 (1.15)	-22.4***
… President of Russia	79.87 (2.21)	84.78 (0.93)	-4.91*

Note: Means are weighted using the full sample weight provided by the ANES and are presented as the percent of respondents who responded correctly to each political knowledge item. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4a: Political Attitudes (2012)

	In-Person	Online	Difference
Democrat	48.51 (1.53)	45.31 (1.06)	3.20
Independent	10.16 (.90)	16.39 (.80)	-6.23***
Republican	41.33 (1.56)	38.30 (1.03)	3.03
Liberal	27.43 (1.52)	25.38 (.93)	2.04
Moderate	30.66 (1.63)	35.97 (1.04)	-5.31**
Conservative	41.91 (1.76)	38.65 (1.04)	3.26
Abortion should always be legal	42.89 *(1.52)*	40.08 *(1.05)*	2.32
Gay marriage should be legal	42.89 *(1.53)*	40.07 *(1.05)*	2.81
Gay couples should be able to adopt	66.18 (1.49)	61.45 (1.04)	4.73**
Gov't should spend less (1-7)	4.14 (.05)	4.34 (.04)	-.19**
Favor private health plan (1-7)	4.04 (.06)	4.26 (.04)	-.22**
Approve of President (1-5)	2.55 (.04)	2.40 (.03)	.15**
Approve of Congress (1-5)	1.77 (.03)	1.74 (.02)	.03

Note: Means are weighted using the full sample weight provided by the ANES and are presented as percentages unless the response scale is included in parentheses. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Party ID is measured on a 7-point scale and partisan leaners are coded as either Democrats or Republicans. Ideology is measured on a 7-point scale and "Haven't thought much about this" and "Don't know" responses are coded as missing values. For policy preference questions, items have been rescored so that higher values denote greater support for the policy position. Approval scores are similarly rescored so that higher values denote greater approval.

Table 4b: Political Attitudes (2016)

	In-Person	Online	Difference
Democrat	45.81 (2.23)	45.8 (1.01)	.01
Independent	18.2 (1.49)	22.26 (0.94)	-4.06*
Republican	43.39 (2.12)	37.89 (1.05)	5.5*
Liberal	30.14 (2.18)	32.04 (1.11)	-1.9
Moderate	26.32 (2.12)	27.45 (0.98)	-1.13
Conservative	43.55 (2.31)	40.51 (1.14)	3.04
Abortion should always be legal	43.2 (2.46)	44.71 (1.06)	-1.51
Gay marriage should be legal	56.35 (1.78)	58.05 (1.05)	1.3
Gay couples should be able to adopt	71.33 (1.98)	71.4 (1.07)	-.07
Gov't should spend less (1-7)	3.92 (0.07)	4.1 (0.4)	-.18*
Favor private health plan (1-7)	3.98 (0.08)	4.06 (0.05)	-.08
Approve of President (1-5)	2.53 (0.06)	2.51 (0.03)	.02
Approve of Congress (1-5)	1.79 (0.04)	1.79 (0.02)	-.01

Note: Means are weighted using the full sample weight provided by the ANES and are presented as percentages unless the response scale is included in parentheses. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Party ID is measured on a 7-point scale and partisan leaners are coded as either Democrats or Republicans. Ideology is measured on a 7-point scale and "Haven't thought much about this" and "Don't know" responses are coded as missing values. For policy preference questions, items have been rescored so that higher values denote greater support for the policy position. Approval scores are similarly rescored so that higher values denote greater approval.

Table 5a. Ideological differences by mode, 2012

	Face-to-Face	Internet	Difference
Issue positions (liberal)	48.09	46.48	-1.61
			(1.01)
Affirmative action (pro)	27.52	28.75	-2.08
			(2.12)
Federal spending (pro)	68.91	58.45	-10.46***
			(0.96)
Environmentalism (pro)	69.62	66.95	-2.68*
			(1.20)
Gay rights (pro)	67.11	62.10	-5.02**
			(1.57)
Abortion (pro-choice)	63.91	64.18	0.26
			(1.65)
Racial resentment (reversed)	38.41	36.10	-2.31
			(1.29)
Modern sexism (reversed)	64.55	62.42	-2.13**
			(0.76)
Egalitarianism (pro)	64.07	60.15	-3.92***
			(0.95)

Note. Standard errors for estimated differences in parentheses. Estimates adjusted for full sample design: weights, clustering, and stratification. All variables normalized to the scale from 0 to 100
* $p < .05$ ** $p < .01$ *** $p < .001$

Table 5b. Ideological differences by mode, 2016

	Face-to-Face	Internet	2016
Issue positions (liberal)	49.74	48.88	-0.86
			(1.13)
Affirmative action (pro)	34.65	34.13	-0.52
			(1.32)
Federal spending (pro)	73.97	66.85	-7.12***
			(0.78)
Environmentalism (pro)	70.52	70.59	0.08
			(1.40)
Gay rights (pro)	56.10	54.36	-1.74
			(1.16)
Abortion (pro-choice)	62.08	63.81	1.73
			(1.97)
Racial resentment (reversed)	45.04	45.30	0.26
			(1.80)
Modern sexism (reversed)	67.04	66.29	-0.75
			(1.10)
Egalitarianism (pro)	49.01	51.24	2.23*
			(0.90)

Note. Standard errors for estimated differences in parentheses. Estimates adjusted for full sample design: weights, clustering, and stratification. All variables normalized to the scale from 0 to 100
* $p < .05$ ** $p < .01$ *** $p < .001$

Table 6a: Presidential Candidate Character Traits (2012)

	In-Person	Online	Difference
Obama Traits (1-5):			
Leadership	3.07 (.04)	2.77 (.03)	.30***
Knowledge	*3.61* *(.04)*	*3.26* *(.03)*	.35***
Empathy	*3.02* *(.04)*	*2.89* *(.03)*	.13*
Honesty	3.04 (.04)	*2.87* *(.03)*	.18***
Romney Traits (1-5):			
Leadership	2.86 (.04)	2.82 (.03)	.05
Knowledge	3.19 (.03)	3.05 (.03)	.14**
Empathy	2.28 (.04)	2.25 (.03)	.03
Honesty	2.58 (.04)	2.55 (.03)	.03

Note: Means are weighted using the full sample weight provided by the ANES. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Respondents were asked to rate whether certain character traits described each presidential candidate on a 5-point scale, ranging from "Extremely well" (5) to "Not well at all" (1).

Table 6b: Presidential Candidate Character Traits (2016)

	In-Person	Online	Difference
Clinton Traits (1-5):			
Leadership	2.76 (0.06)	2.72 (0.03)	.04
Knowledge	3.55 (0.06)	3.37 (0.03)	.17**
Empathy	2.53 (0.06)	2.46 (0.03)	.07
Honesty	2.06 (0.06)	2.07 (0.03)	-.01
Speaks mind	3.04 (0.06)	2.86 (0.03)	.18**
Even-tempered	2.98 (0.06)	2.94 (0.03)	.04
Trump Traits (1-5):			
Leadership	2.61 (0.08)	2.52 (0.03)	.09
Knowledge	2.44 (0.08)	2.28 (0.03)	.17*
Empathy	2.14 (0.08)	2.12 (0.03)	.02
Honesty	2.33 (0.06)	2.26 (0.03)	.07
Speaks mind	4.28 (0.04)	4.14 (0.03)	.14**
Even-tempered	1.83 (0.06)	1.84 (0.03)	-.01

Note: Means are weighted using the full sample weight provided by the ANES. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Respondents were asked to rate whether certain character traits described each presidential candidate on a 5-point scale, ranging from “Extremely well” (5) to “Not well at all” (1).

Table 7: Turnout Model with Mode Interaction (2012)

	Voted in 2012			
	(1)	(2)	(3)	(4)
Female	0.125	0.239**	0.239**	0.244***
	(0.072)	(0.074)	(0.074)	(0.074)
Age	0.037***	0.033***	0.033***	0.032***
	(0.002)	(0.002)	(0.002)	(0.002)
Education	0.469***	0.335***	0.335***	0.333***
	(0.037)	(0.039)	(0.039)	(0.039)
Income	0.202***	0.157***	0.157***	0.153***
	(0.023)	(0.024)	(0.024)	(0.024)
Race: …Black	0.803***	0.830***	0.830***	0.838***
	(0.123)	(0.125)	(0.125)	(0.125)
…Hispanic	0.023	0.115	0.115	0.131
	(0.108)	(0.110)	(0.110)	(0.111)
…Other	-0.499***	-0.497***	-0.497***	-0.487***
	(0.138)	(0.141)	(0.141)	(0.141)
Knowledge		0.471***	0.469***	0.792***
		(0.039)	(0.039)	(0.087)
Online			0.018	0.471***
			(0.077)	(0.128)
Knowledge X Online				-0.406***
				(0.094)
Constant	-2.415***	-2.525***	-2.535***	-2.827***
	(0.165)	(0.168)	(0.174)	(0.188)
Observations	5,258	5,238	5,238	5,238
Log Likelihood	-2,332.562	-2,248.630	-2,248.501	-2,235.550
McFadden's Pseudo R^2	0.162	0.192	0.192	0.196

Note: *p<0.05; **p<0.01; ***p<0.001. Reported are log-odds coefficients. All models are weighted. Education is measured on a 5-point scale: less than hs, hs credential, some post-hs, bachelor's degree, and graduate degree. Race is coded as a factor variable with white as reference. Knowledge is an additive scale representing the number of correct office recognition items (see Table 3a).

Table 8a: Knowledge (2011 Mode Experiment)

	Face-to-Face	Self-Complete	Difference
# Correct Wordsum Items (0-4)	2.55	2.60	-.05
	(.04)	(.05)	
Pluto	71.68	76.44	-4.75
	(2.01)	(1.89)	
Medicine	63.56	67.72	-4.16
	(2.14)	(2.08)	
Moby Dick Author	46.73	57.62	-10.89***
	(2.22)	(2.20)	
Correctly Identified			
…Vice President	56.44	64.55	-8.12**
	(2.21)	*(2.13)*	
…Prime Minister (UK)	20.40	22.38	-1.98
	(1.79)	*(1.86)*	
…Chief Justice	16.24	16.44	-0.20
	(1.64)	*(1.65)*	
…Barack Obama as liberal	50.89	57.82	-6.93*
	(2.23)	(2.20)	
…Mitt Romney as conservative	37.23	40.79	-3.56
	(2.15)	(2.19)	
…Jon Huntsman as conservative	9.90	13.07	-3.17
	(1.33)	(1.50)	

Note: Means are presented as the percent of respondents who responded correctly to each political knowledge item. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. The wordsum is the number of correct responses on a 4-item vocabulary test from the General Social Survey, which asked respondents to identify the correct synonym for the words: space, broaden, allusion, and cloistered. Respondents were also asked whether Pluto was a planet, whether antibiotics kill viruses, and who the author of *Moby Dick* is. Open-ended political knowledge items asked respondents which office John Roberts holds and who holds the office of the Vice President of the U.S. and Prime Minister of the U.K. Finally, respondents were asked to place Barack Obama, Mitt Romney, and John Huntsman on a five-point ideological scale from "very liberal" to "very conservative". Responses were coded as correct if they identified Obama as either "very liberal" or "liberal" were coded as correct, and Huntsman and Romney as either "very conservative" or "conservative".

Table 8b: Favorability Ratings (2011 Mode Experiment)

	Face-to-Face	Self-Complete	Difference
Favorability (1-5)			
Muslims	2.93 (.04)	2.64 (.05)	.29***
Mormon	2.96 (.04)	2.79 (.05)	.17**
Tea Party	2.60 (.07)	2.64 (.07)	-.05
Obama	3.30 (.06)	3.17 (.07)	.13
John Huntsman	2.80 *(.08)*	2.80 *(.07)*	0.00
Mitt Romney	2.81 *(.06)*	2.74 *(.06)*	.06

Note: Respondents were asked whether to rate how favorably they view each individual or group on the following scale: very unfavorable, somewhat unfavorable, neutral, somewhat favorable, or very favorable. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 9a: % "Haven't Though Much About It" (2016)

	Unweighted			Weighted		
Item	In-Person	Online	Diff.	In-Person	Online	Diff.
Ideology	22.44	21.97	0.47*	22.72	19.62	3.05**
Spending & Services	12.45	15.5	-3.05**	11.70	13.98	-2.28*
Defense Spending	11.52	13.98	-2.46*	9.93	12.41	-2.48*
Gov't-Private Healthcare	7.62	12.94	-5.32***	7.79	12.09	-4.30***
Guaranteed Job	9.31	12.01	-2.7**	8.79	10.97	-2.18
Assistance to Blacks	10.16	11.91	-1.75	9.86	10.66	-.79
Environment-Jobs	16.26	16.73	-0.47	16.35	15.14	1.21

Note: Unweighted and weighted proportion of "haven't though much about this" responses.
Source: 2016 ANES.

Table 9b: Likes/Dislikes Battery (2016)

	In-Person	Online	Difference
% Likes something about Democratic candidate for president	50.15 (2.83)	42.22 (1.07)	7.93*
% Dislike something about Democratic candidate for president	66.8 (2.06)	57.2 (1.06)	9.64***
% Like something about Democratic Party	55.58 (2.44)	43.66 (1.09)	11.92***
% Dislike something about Democratic Party	58.6 (2.57)	47.92 (0.99)	10.68***
% Likes something about Republican candidate for president	49.2 (2.45)	39.41 (1.03)	9.79***
% Dislike something about Republican candidate for president	74.33 (2.35)	64.7 (1.12)	9.63***
% Like something about Republican Party	51.7 (2.14)	40.91 (1.0)	10.8***
% Dislike something about Republican Party	65.24 (2.52)	54.0 (1.09)	11.24***

Note: Means are weighted using the full sample weight provided by the ANES. Standard errors are in parentheses. Significance determined using two-tailed independent samples t-tests; * $p < .05$, ** $p < .01$, *** $p < .001$. Respondents were asked to respond yes or no to the above questions with items recoded with "no" responses coded zero. The weighted percentage means reported above are "yes" responses. "Refused" and "Don't Know" responses were coded as missing values.