PA

Paying Attention to Inattentive Survey Respondents

R. Michael Alvarez[1], Lonna Rae Atkeson[2], Ines Levin[3] and Yimeng Li[4]

[1] Professor of Political Science, California Institute of Technology, USA. Email: rma@caltech.edu
[2] Professor of Political Science, University of New Mexico, USA
[3] Assistant Professor of Political Science, University of California, Irvine, USA
[4] Graduate Student, Division of the Humanities and Social Science, California Institute of Technology, USA

Abstract

Does attentiveness matter in survey responses? Do more attentive survey participants give higher quality responses? Using data from a recent online survey that identified inattentive respondents using instructed-response items, we demonstrate that ignoring attentiveness provides a biased portrait of the distribution of critical political attitudes and behavior. We show that this bias occurs in the context of both typical closed-ended questions and in list experiments. Inattentive respondents are common and are more prevalent among the young and less educated. Those who do not pass the trap questions interact with the survey instrument in distinctive ways: they take less time to respond; are more likely to report nonattitudes; and display lower consistency in their reported choices. Inattentiveness does not occur completely at random and failing to properly account for it may lead to inaccurate estimates of the prevalence of key political attitudes and behaviors, of both sensitive and more prosaic nature.

Keywords: survey design, list experiments, survey response, satisficing, instructed-response items, trap questions, sensitive questions

1 Introduction

Over the past two decades the environment in which respondents participate in surveys and polls has changed, with shifts from interviewer-driven to respondent-driven surveying, and from probability to nonprobability sampling. It is still not known precisely how these changes in the survey environment have affected the quality of survey response. Also, response rates for traditional polling have been declining dramatically (Atkeson, Adams, and Alvarez 2014). These changes have focused the attention of survey methodologists on data quality, and on the motivation and engagement of survey respondents. These questions are important in political science where surveys are a primary tool for testing theories of political behavior, and where many researchers use new methodologies like opt-in nonprobability samples with national coverage (e.g. Cooperative Congressional Election Study, CCES), as well as survey respondent workforces such as mTurk or Google Consumer Surveys (for discussions of these survey techniques see Berinsky, Huber, and Lenz 2012; Ansolabehere and Schaffner 2018).

Online panels, and other new technologies such as interactive voice response (IVR), computer assisted personal interviewing (CAPI), and address based sampling (ABS), have been behind

the change from predominantly interviewer-driven survey environments, face-to-face (FTF) and telephone, to respondent-driven ones (Dillman, Smyth, and Christian 2009; Atkeson and Adams 2018). The presence of an interviewer enforces some control over the pace of the interview and social dynamics are believed to increase respondent engagement, while the lack of one gives control of the survey to the respondent, which allows for reduced engagement.

One consequence of these technological changes is that survey respondents in these environments may be less attentive to survey questions. Some respondents may pay little attention to the questions or their responses, while others may deliberately misrepresent their behavior or preferences (Atkeson and Adams 2018). This is a cause for concern, as well-considered responses are necessary for quality survey data. The expectation in a survey environment is that the respondent is mindful in the survey process—reading or listening, and then engaging cognitively to provide a meaningful answer to every survey question. Lack of attentiveness may be a source of nonsampling bias and response error, and a contributor to total survey error (Groves and Lyberg 2010). This may increase the amount of noise in the data, producing inaccurate estimates, and hampering our ability to test hypotheses with precision and accuracy.

Alternatively, noisy data may be inherent in survey research because it may reflect the ambiguity, disinterest, inattentiveness, and distraction that pervades citizen interest in politics and policy (Alvarez and Brehm 2002). In this way, including inattentive respondents in surveys may be important because citizen nonattitudes are prevalent in the public on any particular issue or topic, and accounting for nonattitudes might be crucial for making accurate inferences about research questions. Therefore, it is important to study how to identify engaged and disengaged respondents, and the implications of their responses for testing theories of political behavior.

In many circumstances, simple direct questions may not adequately elicit useful information from respondents. This is particularly true for sensitive issues, where eliciting truthful answers directly is not feasible due to social desirability bias (Maccoby, Maccoby, and Lindzey 1954; Edwards 1957; Fisher 1993). Researchers have developed indirect approaches, such as the randomized response technique (Warner 1965) and the list experiment (Miller 1984), for measuring sensitive behavior and attitudes via opinion surveys (for a recent review see Rosenfeld, Imai, and Shapiro 2016). These techniques involve indirect questions that have longer question wording, more complex structure, and are more cognitively demanding. These types of questions have not been investigated in the past in relation to respondent attentiveness but, given the complexity of popular approaches for measuring sensitive behavior, we expect inattentive respondents to provide much less accurate and consistent information on their sensitive dispositions than those who carefully consider survey questions.

One strategy used in online surveys to detect inattentive respondents is the inclusion of attention checks (i.e. screeners for attention, also called trap or red herring questions), such as instructed-response items and instructional manipulation checks (IMCs), which instruct respondents to answer a question in a specific way.[1] Oppenheimer, Meyvis, and Davidenko (2009) demonstrated that attention checks in the form of IMCs are effective at detecting participants who are not following instructions, increasing statistical power and data reliability. Berinsky, Margolis, and Sances (2014) documented that numerous studies used IMCs from 2006 to 2013, and recommended using multiple IMCs to measure attention. In some research designs, these techniques are used as filters, with failing respondents eliminated from the survey, while in others the information is used to assess data quality.[2] In our study, we use instructed-response items as attention checks to identify inattentive respondents. We examine the characteristics of respondents who failed our trap questions (TQs) and compare them to respondents who did

---

1   We use the terms "attention checks" and "trap questions" interchangeably throughout the paper.
2   Eighty percent of papers documented in Berinsky, Margolis, and Sances (2014) exclude failures from their analyses.

not. We then examine how inattentive and attentive survey respondents answered a series of questions about political attitudes and behavior, including a double-list experiment—an indirect questioning technique introduced in Droitcour *et al.* (1991), which combines two standard list experiments on the same sensitive issue to improve efficiency and gain additional diagnostic opportunities (Glynn 2013).

2 Survey Satisficing and Attention Checks

Although survey researchers want their respondents to be engaged in the survey process, it is likely that some respondents may not be completely engaged. When faced with demanding information-processing tasks some respondents will expend only the minimum amount of effort to provide a response. In psychology Simon (1956) described this as satisficing. In the context of the survey response process, respondents who satisfice may not search their memory completely or may not fully understand the question, and in general they will take a superficial approach to the question–answer format (Krosnick 1991). In extreme cases, respondents may not even pay attention to a question, and engage in random guessing (Oppenheimer, Meyvis, and Davidenko 2009; Berinsky, Margolis, and Sances 2014; Jones, House, and Gao 2015).

A strategy for identifying inattentive survey respondents involves embedding attention checks in carefully selected locations in the survey instrument. One set of attention checks are instructed-response items that ask respondents to provide a specific response and are often part of a grid question or a group of questions with the same scale (e.g., rows in a grid instructing respondents to "please select 'strongly disagree' for data quality control"). Instructed-response items evaluate respondents' compliance with simple and concise instructions and have been used in several previous studies as attention indicators (Barber, Barnes, and Carlson 2013; Ward and Pond 2015; Bowling *et al.* 2016). Some inattentive respondents may pass the attention check by chance, but this problem can be mitigated with the inclusion of multiple attention checks. A closely related technique involves infrequency scales or bogus items. These are items on which all or virtually all attentive respondents should provide the same response, e.g. "I have never used a computer" (Huang *et al.* 2012). In their research, Meade and Craig (2012, p. 16) conclude "we strongly endorse bogus items—or, preferably, instructed response items (e.g. "Respond with 'strongly agree' for this item")—for longer survey measures."

Another popular type of attention check is the IMC first introduced by Oppenheimer, Meyvis, and Davidenko (2009). A typical IMC evaluates whether respondents read and follow instructions within a lengthy question prompt—instructions taking precedence over other requests for information made elsewhere in the question's text (e.g. may include phrases such as "please ignore the rest of the question and select options A and C"). In contrast to instructed-response items, which would not make sense if respondents ignore the part of the question requesting specific responses, IMCs are perfectly valid questions without the phrase containing the instruction, and could therefore be more easily misunderstood by inattentive respondents. Both type of attention checks may be used for general inattention detection, but IMCs are particularly useful for survey experiments where "the manipulation of a study is hidden in similar text somewhere else in that study" (Curran 2016, p. 14) and instructed-response items are particularly useful for grid and check-all-that-apply questions.

Besides attention checks, previous research suggests that indices capturing response patterns, such as overall response duration and frequency of nonattitudes, may provide valuable information about respondent attentiveness (Johnson 2005; Huang *et al.* 2012; Meade and Craig 2012; Maniaci and Rogge 2014; Curran 2016). We consider consistency of expressed preferences, propensity to select the same response to multiple contiguous questions (i.e. straightlining behavior), and response time as additional validation for attention measured based on multiple instructed-response items.

Depending on the reason for the failure to pass attention checks, trapping and removing inattentive respondents from the sample may or may not be a reasonable approach for dealing with respondent satisficing (Downes-Le Guin 2005; Zagorsky and Rhoton 2008; Berinsky, Margolis, and Sances 2014). If the motives behind satisficing behavior correlate with respondent characteristics influencing the outcome of interest, then listwise deletion of inattentives could lead to inaccurate inferences as inattentive responses would not be *missing completely at random* (MCAR; Little 1992). Past research suggests that respondents who fail TQs are younger (Kapelner and Chandler 2010; Berinsky, Margolis, and Sances 2014), more likely to be male (Kapelner and Chandler 2010; Berinsky, Margolis, and Sances 2014), less educated (Kapelner and Chandler 2010), and more likely to be non-White (Berinsky, Margolis, and Sances 2014, but see Anduiza and Galais 2016). These findings are partially consistent with our results and suggest that caution should be exercised in deciding whether to keep or drop inattentives. If, unobserved factors influence both attentiveness and the mechanism behind the outcome of interest, then it would also be inappropriate to treat inattentive responses as *missing at random* (MAR) conditional on covariates. Missingness resulting from the process of removing inattentives would then be *missing not at random* (MNAR) and therefore *nonignorable*, and it may not be possible to successfully address it via conventional model-based imputation procedures (King *et al.* 2001; Pepinsky 2018).

Ultimately, as we discuss toward the end of the paper, how to deal with inattentiveness comes down to a comparison of costs and benefits associated with keeping inattentive respondents in the sample. Designing a survey where attention checks are used, and both attentive and inattentive survey respondents complete the survey, produces noisy data at significant cost. In that context, statistical adjustments may become necessary to account for differences in attentiveness between respondents. On the other hand, designing a survey where inattentive survey respondents are dropped from the sample lowers survey costs considerably (fewer respondents need to be interviewed for the entire survey), reduces noise, but risks producing an unrepresentative sample that will require post hoc statistical adjustment to produce population-level inferences.

3 Data and Methods

3.1 Data

We use data from an online survey of 2,725 California adults conducted in July 18–30, 2014 using a sample recruited by Qualtrics through the e-Rewards panel.[3] Recruitment into the e-Rewards online panel was carried out using a double opt-in process, whereby "[a]fter receiving a personalized email invitation to join the e-Rewards program, individuals must opt-in and agree to provide truthful and well-considered answers [. . .]. After the first opt-in during the enrollment process, the individual is sent a follow-up e-mail confirmation that requests for him/her to click on a link to validate opt-in. [. . .] Once a member has completed the double opt-in process, they are then eligible to begin receiving survey invitations" (e-Rewards 2008, p. 3).

In this survey panel participants were invited via e-mail to complete a 20-minute respondent-driven online questionnaire, which was designed and implemented using Qualtrics survey software.[4] Respondents could choose between English and Spanish versions of the questionnaire.[5] The data collection process began on July 18, 2014 and concluded on July 30, 2014 when the target sample size of 1,700 complete and attentive responses was reached. Individuals that failed to meet gender, age, and education quotas were filtered out at the beginning of the survey after completing

---

3 The sample of 2,725 respondents includes 1,750 responses from individuals who completed the entire questionnaire and 975 partial responses from individuals who were filtered out before the end of the survey after failing a TQ. Data and code for replicating the analyses reported in this paper are available (Alvarez *et al.* 2018).

4 Figure A1 in Appendix A shows a screenshot of a sample e-mail invite.

5 Only 21 respondents (less than 1%) chose to see the survey in Spanish.

a brief screener section assessing basic demographics.[6] Respondents who reached the end of the survey answered additional demographic questions that were later used to construct survey weights. Those weights are not used in this paper as the demographic data used in calculating weights are only available for respondents who completed the entire survey—i.e. are not available for respondents that failed TQs. Estimates reported in this paper apply to the sample at hand, and do not necessarily reflect the California adult population.

3.2 Identification of inattentive respondents

Research on attention checks suggests that many respondents within any research design may be inattentive and hence be more likely to satisfice. Estimates suggest the range of failure for attention checks is quite large, from as low as 8% to 50% of respondents (Miller 2006).

In our study, we had three instructed-response items that appeared at different points in the survey and used different means to assess whether a subject was paying close attention to the survey questions. Figures A2–A4 in Appendix A show screenshots of desktop versions of these questions. The three TQs were:

- TQ 1: A stand-alone question located immediately below a check-all-that-apply question on participation in twelve political activities. It instructed respondents to type the word "government" inside a text box. Answers were coded as correct if they contained the term "gov."
- TQ 2: Grid question that asked respondents to report support or opposition to several policies. In one of the rows, respondents were instructed to select "I'm indifferent" for quality purposes. The location in the grid of the row containing the instructed-response item was randomly determined.
- TQ 3: Grid question that asked respondents to agree/disagree with statements indicating varying levels of tolerance toward other people's views. In one of the rows, respondents were instructed to select "Disagree Strongly" for quality purposes. The location in the grid of the row containing the instructed-response item was kept fixed at the bottom.

Respondents that failed each instructed-response item were filtered out. Nonetheless, incomplete responses provided by respondents who eventually failed a TQ were recorded up to the moment they were dropped from the survey. This allowed us to use responses to questions preceding TQs to compare the attributes of those who did and did not survive each attention check, and subsequently evaluate how inattentiveness distorts the observed distribution of political attitudes and behavior. While inattentive respondents may have passed some of these attention checks by chance—particularly the ones embedded within a grid—the chance of surviving all attention checks should be considerably higher for attentive respondents than for those engaging in satisficing behavior. In the rest of this article, we refer to respondents that passed all checks as *attentives*, and to the rest as *inattentives*.

In the next section, we provide some basic information that profiles the characteristics of attentive and inattentive survey respondents. We then validate our attentive–inattentive distinction by considering alternative measures of response quality (response time, item nonresponse, and response consistency) for attentives and inattentives. We then provide evidence on systematic differences between attentive and inattentive respondents in terms of their reported political attitudes and behavior.

---

6 Survey quotas were selected as to ensure a minimal number of complete responses within predefined gender, age, and education categories. To ensure the timely completion of the data collection process, we allowed for overrepresentation of age/education groups. Quotas were removed on July 29, 2014, after more than 1,600 complete responses had been collected, to speed up the conclusion of the data collection process.

4   Results

In total, slightly more than one third (36%) of respondents were inattentive, as they failed to pass one of the TQs (see Table B1 in the supplementary materials).[7] The large incidence of failure to pass attention checks suggests that inattentiveness is a common phenomenon in respondent-driven surveys like this one.

Among respondents that failed the first attention check, the most common behavior was leaving the text box empty and skipping the question (displayed by 96% of those that failed TQ 1). Among respondents that failed the second attention check (which instructed individuals to select "I'm indifferent" along a scale that also included the options "I don't know," "Oppose," and "Support"), the most common behavior was selecting "Support" (selected by 44% of those that failed TQ 2, and coinciding with the modal response to all other items included in the grid). We did not find that the order of the instructed-response item in the grid, which was randomized for the second attention check, had a significant influence on failure rates. It is possible that some inattentives passed TQ 2, as respondents were instructed to report a nonattitude ("I'm indifferent"). This was not the case for TQ 3, where the instructed response represented a definite stance ("Disagree Strongly") and was an uncommon answer to other items in the grid. Among respondents that failed the third attention check, 60% selected "Neither Agree nor Disagree," a nonattitude, in response to TQ 3.

In addition to being commonplace, inattentiveness does not occur completely at random. Those who passed all attention checks differ demographically from those that failed (see Table B2 in the supplementary materials). Less educated and younger respondents are more likely to fail, consistent with Kapelner and Chandler (2010) and Berinsky, Margolis, and Sances (2014), but we find no gender differences in attentiveness, which contrasts with previous studies. We also find that inattentives interact with the survey in specific ways: they answer questions more quickly, consistent with Oppenheimer, Meyvis, and Davidenko (2009); are more likely to answer "don't know"; and are more likely to straightline (see Table B3 in the supplementary materials). These latter two results are consistent with past work on satisficing by survey mode (see Atkeson and Adams 2018).

Are the three TQs similar in terms of their ability to detect inattentives? Since respondents that failed an attention check were filtered out, we are not able to compare the characteristics and behavior of respondents who fail each TQ but pass the rest. We can, nonetheless, learn much about the operation of each TQ by evaluating whether two reasonable expectations hold within the following nested groups of respondents: those that see TQ 1 (the entire sample), those that see TQ 2 (including only respondents that pass TQ 1), and those that see TQ 3 (including only respondents that pass TQ 2).

Within each of these groups, we examine whether respondents that pass the next TQ look and behave differently than those that fail. Because of differences in the composition of these three groups, however, differences between those that fail and pass each check cannot be compared between groups. But given that individuals who pass more checks are likely to be more attentive than individuals who pass fewer checks, we expect that a typical respondent surviving all checks, for instance, will not be less attentive than the typical respondent in the entire sample. Conditional on this testable assumption (which can be validated based on auxiliary information on attentiveness collected for all respondents before exposure to TQ 1, such as response speed up to that point), if the three TQs operate similarly, then observed differences between respondents that pass and fail each check should become progressively smaller as we move from considering all respondents to only considering those that get to see TQ 3, as the most inattentive respondents

---

7 The breakdown of the failure rates over the three TQs was as follows: 21%, 8%, and 6% of respondents were filtered out after failing to pass the first, second, and third attention checks, respectively.

PA

will have been filtered out before reaching the last attention check. Observed differences between respondents who pass and fail TQ 1, for instance, should be more pronounced than differences between respondents who pass and fail TQ 3, as respondents that get to see TQ 3 should display higher (and more homogeneous) levels of attentiveness, on average, than the broader group of respondents including individuals that eventually failed TQ 1 or TQ 2.

Using the information provided in Tables 2 and 3, we find that younger and less educated respondents are more likely to fail TQ 2 and 3. Respondents who fail these two attention checks, in turn, spend less time considering survey questions, report higher rates of nonattitudes, and are more likely to display incomplete and/or intransitive preferences over policy options. These patterns, however, are less pronounced in the case of TQ 1. Differences between respondents that pass and fail TQ 1—in terms of characteristics listed in Table B2 and behaviors listed in Table B3—are no larger than differences between respondents that pass and fail TQ 2 and 3, respectively, a finding that contradicts our second expectation.

These results indicate that TQ 1 (an open-ended instructed-response item that instructed respondents to type the word "government" in a text box, located immediately below a check-all-that-apply question with numerous response alternatives) operates differently than TQs 2 and 3 (instructions given in rows of two separate grid questions, which instructed respondents to select specific responses along a labeled scale). While we cannot evaluate why different types of TQs filter out different respondents, we argue that TQ 1 filtered out respondents for reasons having little to do with inattentiveness, which would explain why respondents that fail TQ 1 differ little in terms of demographic attributes and earlier interaction with the survey instrument relative to those that pass TQ 1. Unlike instructions given in TQ 2 and 3, instructions given in TQ 1 did not state that the request made in the question was for quality purposes. It is possible that many respondents found TQ 1 senseless and as a result decided to ignore the request. Another possibility is that some respondents may have failed to notice the text-box question, as it was located immediately below a check-all-that-apply question with numerous response alternatives (see Figure A2 in Appendix A), rather than in a stand-alone page (as most other questions in the survey). Both of these explanations help account for the observation made before that most respondents that failed TQ 1 did so not because they wrote something other than "government" in the text box, but because they left the text box empty.

In sum, we find that respondents that pass all attention checks differ from those that fail, but these differences are more pronounced for closed-ended grid-style instructed-response items than for the open-ended text-box-style instructed-response item. This suggests that some TQs may be more reliable than others and could make a big difference in terms of who gets filtered out or flagged as inattentive, as could slight differences in question wording (such as whether the question is designated with the intent of verifying response quality or whether it is displayed in the same page as other survey questions). We leave further exploration of this question for future research.

4.1 Direct questioning techniques

To consider the extent to which attentives and inattentives provide different answers to direct questions, we examine their responses to closed-ended questions about their political knowledge, self-reported political activity, and scale placement on policy issues. These measures may demonstrate whether political interest is associated with attentiveness.

We used responses to four questions about factual knowledge of California politics to construct a 0–4 political knowledge scale. The first two knowledge questions asked about the majority party in the State Senate and Assembly. The other two asked about majorities required for passing constitutional amendments and for raising taxes. Figure 1 shows the distribution of the political knowledge scale for attentive and inattentive respondents. These results suggest that inattentives
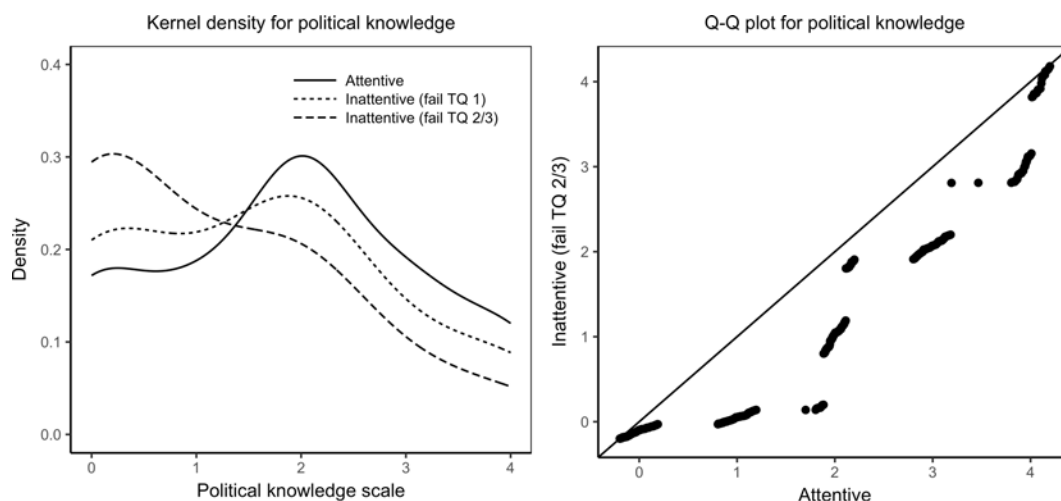
Note: The plot on the left displays the kernel distribution of political knowledge (measured on a 0–4 scale) among respondents that passed all attention checks (solid line), failed the first check (dotted line), and failed the second or third check (dashed line). The quantile–quantile plot on the right compares quantiles of the distribution of political knowledge between respondents that passed all checks (horizontal axis) and those that failed the second or third check (vertical axis).

**Figure 1.** Attentiveness and political knowledge.

miss factual knowledge questions more frequently. In the case of the two questions on majority requirements, inattentives are about as likely as attentives to select "I don't know" (34% and 37% of the time for each question, respectively, the same as attentives). More inattentives fail these knowledge questions because they are more likely to select the wrong percentage (in particular, 6% of inattentives report that unanimity is required for each decision to pass, compared to only about 2% of attentives). In the case of the knowledge items asking about majority party in the State Senate and Assembly, inattentives are both considerably more likely to report "I don't know" and to get the answer wrong by selecting "Republican."[8]

We then used responses to a check-all-that-apply question about participation in twelve political activities to construct a 0–12 political participation scale that was asked immediately before exposure to the first attention check (for a screenshot, see Figure A2 in the Appendix). Listed activities included voting in national and statewide elections, other conventional forms of involvement, and involvement in unconventional acts. Figure 2 shows the distribution of the political participation scale for attentive and inattentive respondents. Summary statistics of self-reported participation in each activity are indicative of nonrandom selection of a small number of response alternatives by inattentives, rather than entirely haphazard choices, as these respondents were consistently more likely to select common conventional activities (e.g. voting and signing petitions), than more demanding activities (e.g. working for campaigns, attending political meetings, and donating) or unconventional ones (e.g. participation in protests and sit-ins). These results suggest that inattentives report participating in fewer political activities due to lower levels of political engagement compared to attentives.

Last, we used responses to six questions on support for liberal policies to construct a 13-point ideology scale (ranging from −6 to 6). Respondents were asked about support for the Affordable Care Act, repealing "Don't Ask, Don't Tell," providing a path to legal status and citizenship for undocumented immigrants, implementing stricter carbon emission limits, restricting the sale

---

8 In the question about the majority party in the State Senate, 29% of inattentives report "I don't know" and 21% incorrectly select "Republican," compared to 20% and 14% of attentives, respectively. In the question about the majority party in the State Assembly, 35% of inattentives report "I don't know" and 20% incorrectly select "Republican," compared to 29% and 10% of attentives, respectively.
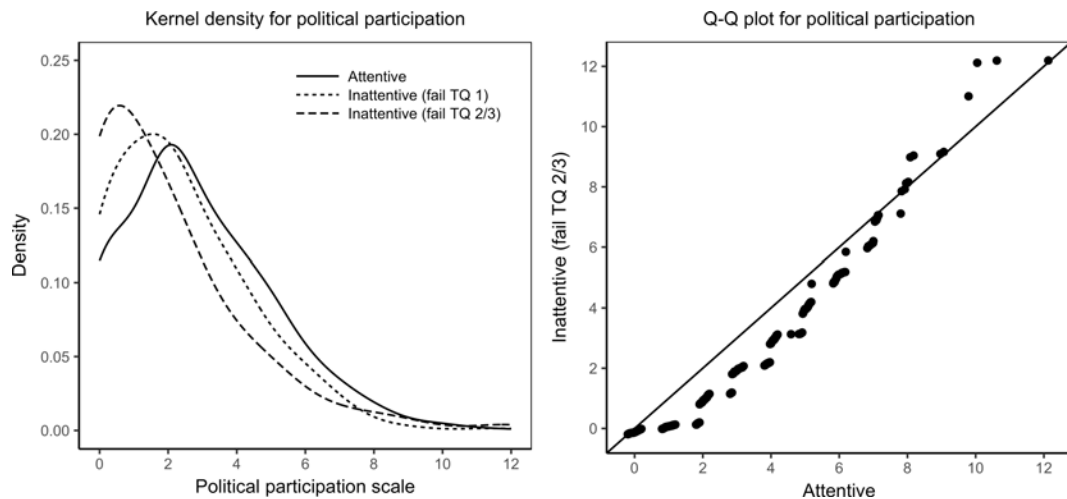
Note: The plot on the left displays the kernel distribution of political participation (measured on a 0–12 scale) among respondents that passed all attention checks (solid line), failed the first check (dotted line), and failed the second or third check (dashed line). The quantile–quantile plot on the right compares quantiles of the distribution of political participation between respondents that passed all checks (horizontal axis) and those that failed the second or third check (vertical axis).

Figure 2. Attentiveness and political participation.

of semiautomatic and automatic weapons, and limiting NSA's collection of domestic phone records. We coded support for each policy as 1 for respondents selecting "support," −1 for those selecting "oppose," and 0 for those selecting "I'm indifferent" or "I don't know." Figure 3 shows the distribution of the ideology scale, constructed by adding up support across the six policies, for attentive and inattentive respondents. These results indicate that inattentives report less conclusive stances toward policy issues. When looking at responses issue by issue, it is evident that inattentives do not select "I don't know" or "I'm indifferent" with equal probability for each policy issue. Among inattentives that select a nonattitude, we find that they are about twice or more as likely to select "I'm indifferent" than "I don't know" (particularly for repealing "Don't Ask Don't Tell" and the gun control item, where they are close to three times as likely to select "I'm indifferent" instead of "I don't know"; and with the exception of the health care law, where they are about as likely to select "I'm indifferent" as "I don't know"). These item-by-item findings suggest that the lower-intensity answers reported by inattentives reflect greater uncertainty about policy positions.

We estimated a series of linear regression models to evaluate whether attitudinal and behavioral differences between attentives and inattentives persist after controlling for demographic backgrounds of respondents that pass and fail attention checks (see Tables B4–B6 in the supplementary materials). The results of these analyses are consistent with those presented before. When inattentives are asked closed-ended questions about political knowledge, civic engagement, and opinions on policy issues, their responses reveal less knowledge, lower involvement, and weaker policy stances than attentives. These results do not allow establishing whether inattentives are shirkers who check fewer activity boxes and provide hasty responses, whether inattentiveness springs from genuine lack of interest in politics, or whether a mixture of both mechanisms is at play. A consideration of responses to indirect questioning may help to clarify the implications of these results for understanding inattentives.

4.2 Indirect questioning techniques

We have looked at direct questions on respondents' political knowledge, participation, and ideology placement. We now evaluate respondents' attitudes toward immigration through a
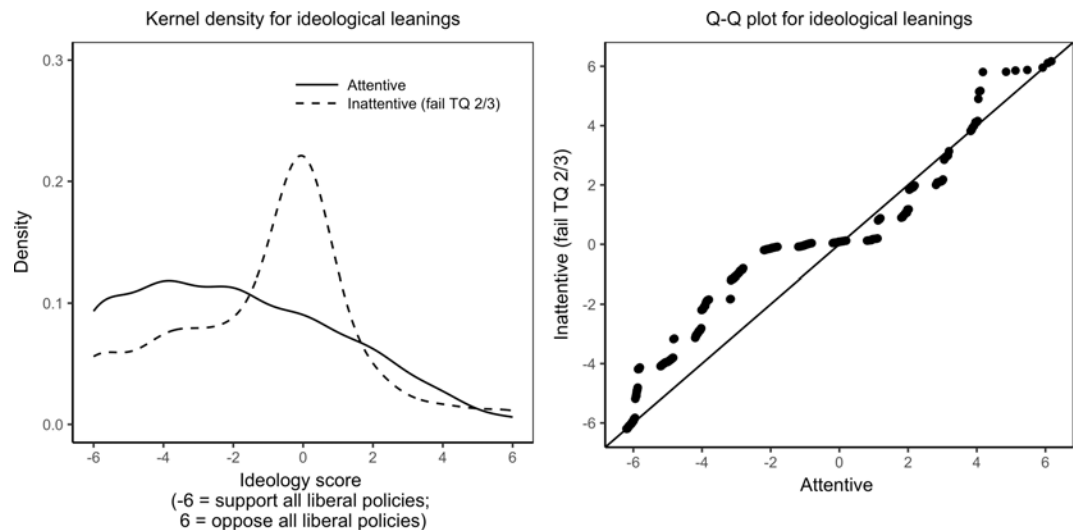
Note: The plot on the left displays the kernel distribution of ideological leanings (measured on a -6 to 6 scale) among respondents that passed all attention checks (solid line) and those that failed the second or third check (dashed line). The quantile–quantile plot on the right compares quantiles of the distribution of ideology between respondents that passed all checks (horizontal axis) and those that failed the second or third check (vertical axis).

**Figure 3.** Attentiveness and ideological leanings.

double-list experiment embedded in the survey.[9] The experiment was designed to measure Californians' support for two state chapters of national anti-immigrant organizations. To preserve the anonymity of these organizations, we refer to them as Organization X and Organization Y. Organization X was described in the double-list experiment as "advocating for immigration reduction and measures against undocumented immigration," and organization Y as a "citizen border patrol group combating undocumented immigration."

The double-list experiment comprised two questions containing "list A" and "list B." The first question (list A) exposed respondents to a list of different groups and organizations in randomized order, and asked them to specify "how many of these groups and organizations you broadly support."[10] This was followed by a second list of different organizations (list B). The two questions provided the name and a brief description of all listed organizations, and were located after the first attention check, but before the second one.

Depending on their treatment status in the double-list experiment, respondents saw different versions of list A and B that included or excluded the name and description of one of the two anti-immigrant organizations. Respondents assigned to the "control A–treatment B" condition were exposed to the sensitive item in list B; and those assigned to the "treatment A–control B" condition were exposed to the sensitive item in list A. Sensitive items displayed under either of these two experimental conditions were randomly assigned to respondents. Items displayed

---

9 In a standard list experiment, respondents in the control group see a list of control items and respondents in the treatment group see a similar list that also includes a sensitive item. A double-list experiment consists of two lists presented to respondents, with different control items but the same sensitive item for respondents seeing the "treatment" version of each list. Respondents are randomly assigned to treatment (i.e. seeing the sensitive item) in the first or second list. Thus, in contrast to a standard list experiment, where only a subset of respondents (those in the treatment group) is exposed to the sensitive item, all respondents in a double-list experiment see the sensitive item at some point (either in the first or second list), leading to potential efficiency gains.

10 The exact instructions given to respondents were: "*Below is a list with the names of different groups and organizations on it. After reading the entire list, we'd like you to tell us how many of these groups and organizations you broadly support, meaning that you generally agree with the principles and goals of the group or organization. Please don't tell us which ones you generally agree with; ONLY TELL US HOW MANY groups or organizations you broadly support. HOW MANY, if any, of these groups and organizations do you broadly support.*"
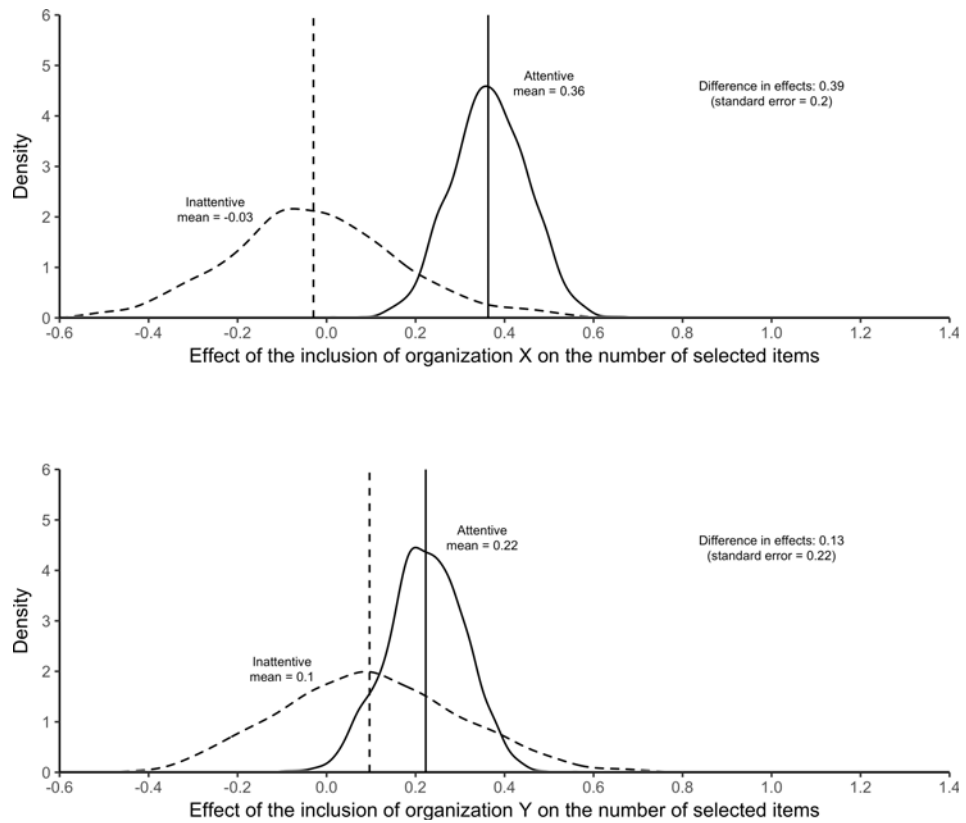
Note: The figure displays the distribution of difference-in-means estimates (mean response under treatment condition minus mean response under control condition) for attentive and inattentive respondents separately in 1,000 bootstrapped samples. Solid and dashed vertical lines show the difference-in-means estimates in the original sample for attentives and inattentives, respectively. The top-right figure in each plot reports the attentive–inattentive difference in difference-in-means estimates together with the corresponding bootstrapped standard error. Results for organizations X and Y are shown on the top and bottom plots, respectively.

Figure 4. Attentiveness and difference-in-means estimates (list A).

to respondents and the number of respondents assigned to each combination of experimental conditions and sensitive items are shown in Tables B7 and B8 in the supplementary materials, respectively.

Possible responses to each list question comprised integers between zero and four under control and between zero and five under treatment (X or Y), representing the number of supported organizations. Under two assumptions—no design effect and no liars—the difference between the average response under treatment and control consistently estimates the level of support for the sensitive item (Imai 2011; Blair and Imai 2012).

Figure 4 provides a visualization of difference-in-means estimates for attentive and inattentive respondents, using list A responses. In the figure the vertical lines show the difference-in-means estimates in our sample and the curves show the distribution of these estimates in 1,000 bootstrapped samples. Attentives select 0.36 more items on average under the X-treatment (and 0.22 more items on average under the Y-treatment) than under control, whereas inattentives select a similar number of items under treatment and control. The distributions for inattentive respondents are more dispersed due to noisier responses and smaller sample sizes.[11]

---

11 This difference in dispersion is confirmed by a Levene's test (the F statistic from this test is 437 for organization X and 419 for organization Y).

Inattentive respondents on average choose fewer items under both control and treatment, with the difference more pronounced under treatment (see Table B9 in the supplementary materials). The decrease is consistent with two types of survey satisficing. First, not supporting a group can be an expression of a nonattitude. As we saw from previous sections, inattentive respondents are more likely to report attitudes in the middle of the scale as opposed to the extremes consistent with both shirking and disinterest. However, the list experiment suggests that shirking may be a primary factor. Alternatively, it could be that many inattentives did not pay attention to the list and chose a small number, especially the first option—zero in this case (see Table B10 in the supplementary materials for the distribution of responses).

This result suggests that inattention may account for artificial deflation due to list length documented in the literature, namely that estimators are biased due to the different list lengths provided to control and treatment (Kiewiet de Jonge and Nickerson 2014, 659).[12] Kiewiet de Jonge and Nickerson (2014) find significant underestimation of the occurrence of a common behavior. In a recent paper, Eady (2017) included a screener for a large-scale list experiment ($n = 24{,}020$) and excluded respondents who failed the screener from the analysis. We reanalyzed data extracted from Eady's replication package (Eady 2016) and found that respondents who failed the screener on average chose a smaller number of items under both control and treatment, with the difference more pronounced under treatment (see Table B12 in the supplementary materials). The difference-in-means estimate for those who passed the screener is 0.88, and the estimate for those who failed is 0.75 (difference: 0.13, s.e. = 0.05).[13]

In our double-list experiment, results are dramatically different for list B, as shown in Figure 5. Attentives select 0.25 more items on average under the X-treatment (and 0.28 more items on average under the Y-treatment) than under control, which is very similar (and nearly identical) to what we showed in Figure 4 for the attentives. However, differences in means for the inattentives are now positive and large in magnitude (0.60 for organization X and 0.53 for organization Y).[14] The cross-list difference for inattentives in terms of the difference-in-means estimates is 0.62 for organization X (s.e. = 0.34) and 0.43 for organization Y (s.e. = 0.35).
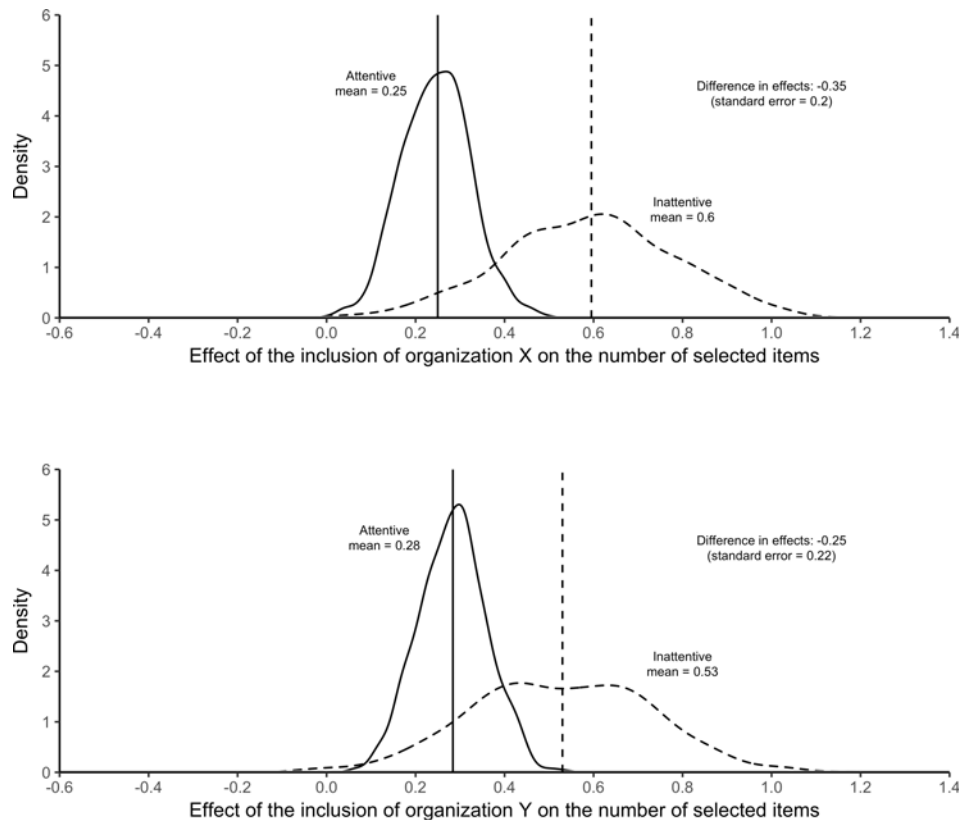
To see why the difference-in-means estimate for inattentive respondents is small for list A (Figure 4), but large for list B (Figure 5), notice the structure of the double-list experiment is such that treated respondents for list A are now under control, and control respondents for list A now receive treatment. In the crosstabs shown in Table B11 in the supplementary materials, we see that many respondents choosing 0 in list A continue to choose 0 in list B; also, those choosing the maximal number continue to choose the maximal number. This tendency is much more pronounced for inattentive respondents, and we interpret this as a type of anchoring (picking the same number for the second list as for the first list). Given that inattentives choose 0 too often under the treatment condition (relative to under control) in list A and exactly the same respondents receive control in list B, they choose 0 too often now under control (relative to under treatment). Also, inattentive respondents under control choose the maximal number more often in list A and now they get the treatment condition. The anchoring leads them to choose the maximal number too often now under the treatment. These are the mechanics behind the reversal seen between Figures 4 and 5.

Our findings have two important implications. First, our results indicate that responses from the second list in a double-list experiment may be constrained by an "anchoring effect."

---

12 Artificial deflation due to length effect can also arise if the inclusion of the sensitive item provides a strong contrast that reduces the attractiveness of control items on the list. We view the two causes as complementary and the magnitude attributable to each cause depends on survey environment as well as items on the list.

13 In our list experiment, an affirmative latent response to the sensitive item is sensitive, whereas it is the opposite for Eady (2017). This provides further evidence that the deflation is artificial and has nothing to do with social desirability bias.

14 Again, the distributions for inattentive respondents are more dispersed, confirmed by a Levene's test (the F statistic is 502 for organization X and 515 for organization Y).

Note: The figure displays the distribution of difference-in-means estimates (mean response under treatment condition minus mean response under control condition) for attentive and inattentive respondents separately in 1,000 bootstrapped samples. Solid and dashed vertical lines show the difference-in-means estimates in the original sample for attentives and inattentives respectively. The top-right figure in each plot reports the attentive–inattentive difference in difference-in-means estimates together with the corresponding bootstrapped standard error. Results for organizations X and Y are shown on the top and bottom plots, respectively.

Figure 5. Attentiveness and difference-in-means estimates (list B).

This observation implies that inattentiveness may undermine the premises of a double-list experiment, overshadowing any efficiency gain. Second, as none or the maximal number of items are chosen disproportionately too often or rarely under different conditions for inattentive respondents, we have identified a violation of the assumption of no design effect. Survey inattention may result in violation of a key assumption underlying the experimental design and may thus undermine the value of these designs.

Since attentive and inattentive respondents also differ demographically, we estimated linear regression models to evaluate whether differential responses persist after controlling for basic individual attributes (see Table B13 in the supplementary materials). According to our most comprehensive specification (Model 3), inattentive respondents on average choose 0.28 fewer items than attentive respondents, holding demographics and other individual characteristics fixed. Respondents who failed either TQ 2 or TQ 3 are much less likely to support anti-immigration groups according to list A (by 39% and 11%, respectively, in absolute magnitude), and much more likely to support anti-immigration groups according to list B (by 44% and 33%, respectively).[15] All effects are large in magnitude and are statistically significant except for the 11% decrease for organization Y for list A. The reversal seen here across lists represents strong evidence that inattentiveness is driven more by shirking than by genuine disinterest in politics.

---

15 44% and 33% come from the following calculations: $0.44 = 0.83 - 0.39$, $0.33 = 0.44 - 0.11$.

5 Discussion: Dealing with Inattentiveness

As we have shown, most polls and surveys are likely to contain many inattentive respondents, and they provide different responses to direct and indirect survey questions from attentive respondents. What should we do to deal with survey inattentiveness? Four general approaches for addressing inattentiveness include: (1) doing nothing (keeping all respondents in the sample and ignoring attentiveness in data analyses); (2) dropping respondents flagged as inattentive and analyzing the rest of the data without further adjustment; (3) dropping respondents flagged as inattentive and reweighting the rest of the data; and (4) keeping all respondents in the sample and accounting for attentiveness via model-based statistical adjustment.

If lack of careful attention to the questionnaire is innocuous, in the sense that it does not alter responses to survey questions, then the best approach for dealing with inattentiveness is to do nothing. This is reasonable in situations where inattentiveness is associated with a lack of interest in politics, provided that uninterested respondents answer similarly regardless of the amount of attention to the questionnaire and concentration effort. In such cases, behaviors such as selecting "don't know" or indicating indifference between response alternatives constitute genuine reflections of inattentives' attitudes toward politics, and should therefore not require adjustment.

If inattentives report different answers than they would were they paying attention—as could be the case for respondents that engage in satisficing behavior for reasons other than lack of interest in politics—then doing nothing is not reasonable, as it could lead to inaccurate inferences. Oppenheimer, Meyvis, and Davidenko (2009) instead argue that "eliminating participants who are answering randomly ... will increase the signal-to-noise ratio, and in turn increase statistical power." Dropping inattentives, the second option, improves efficiency by reducing noise. However, depending on the study and particularly the subject pool, attention may be correlated with individual characteristics, such as age, gender, and education. This is not the case for Oppenheimer, Meyvis, and Davidenko (2009) but is the case for Berinsky, Margolis, and Sances (2014), and our study. For the latter case, if these measured and unmeasured individual characteristics are important correlates of the outcome of interest, then simple elimination of inattentive respondents would lead to a sample that is not representative of the target population (Berinsky, Margolis, and Sances 2014).

An alternative approach in these situations is to drop inattentive respondents from the analysis and reweight the sample to obtain population estimates. If the weighting scheme adequately accounts for the probability of inclusion in the sample being inversely related to correlates of inattentiveness—such as being young and having low levels of education—then this approach could help correct for sampling bias. The performance of reweighting depends on what inattentives' responses would be, were they to pay attention. If these counterfactual responses are close to those given by attentives with similar individual characteristics, then this approach recovers the true population quantities of interest. By training inattentive respondents, Oppenheimer, Meyvis, and Davidenko (2009) find that forcing respondents who fail IMCs to try again until they pass converts inattentive to attentive respondents. However, Berinsky, Margolis, and Sances (2016) were unable to replicate this finding suggesting that more research is necessary to determine inattentives' counterfactual responses.

Dropping inattentives and reweighting the sample, however, also presents a few limitations. First, it assumes that analysts have access to valid and reliable measures of attentiveness that they can use in deciding whether to keep or drop respondents. Since respondents' concentration efforts are not directly observed, attentiveness is likely to be measured with error. This is supported by Berinsky, Margolis, and Sances's (2014) work that finds that there is not a perfect correlation across IMCs both within and across surveys. Moreover, polar cases of complete absence or presence of attention are likely to be rare; therefore, deciding the minimum level

of attention necessary for keeping respondents in the sample may be far from straightforward. Second, inattentiveness may depend on individual characteristics associated with the outcome of interest. In that case, dropping inattentives from the sample may alter estimates, a problem analogous to using listwise deletion for handling cases of data *missing not at random* (Pepinsky 2018). For example, if inattentives genuinely have more moderate opinions on policy issues than attentives, then dropping inattentives could lead analysts to infer that public opinion is more polarized than it actually is (Alvarez and Franklin 1994; Alvarez 1997). Third, dropping inattentives would exacerbate existing unit nonresponse problems and require analysts to rely more heavily on survey weights than they otherwise would. Adjusting survey weights to account for unit nonresponse due to inattentiveness would require access to auxiliary variables predictive of inclusion in the sample (i.e. of propensity to provide attentive responses), information that may not always be observable or available to practitioners (Bailey 2017). And last, this approach does not allow evaluating the ways inattentiveness distorts answers to survey questions, which might be of substantive interest to some researchers.

A fourth approach is to develop a statistical model relating outcome variables to measure(s) of attentiveness, controlling for individual attributes that may be associated with both the attention paid to the questionnaire and the attitude or behavior of interest. In the absence of systematic error in the measure of attentiveness, this model-based approach could be used to learn about the relationship between inattentiveness and expected values of the outcome variable. When multiple indicators of attentiveness are available in the data set, analysts may be able to incorporate information about measurement error associated with attention assessments into the analysis, which would lead to more accurate estimates of uncertainty about quantities of interest (e.g. standard errors accounting for uncertainty in the attention assessment).

Ultimately, researchers must weigh the benefits of measuring respondents' attention to the questionnaire and adjusting for inattentiveness, against the costs of doing so. In the case of the survey analyzed in this paper, for instance, the polling firm recommended the inclusion of attention filters and did not charge for incomplete responses from respondents that failed TQs. Because of budget constraints, a decision was made to follow this advice and filter out inattentive respondents. Collecting measures of attentiveness via attention checks also requires lengthier questionnaires and increased administration times. It may also have other consequences including the inducement of Hawthorne effects by motivating participants to provide socially desirable responses or to censor their responses because of fears that anonymity has been lost (Clifford and Jerit 2015; Vannette 2017). But not including attention checks (or failing to collect auxiliary information on attention) prevents the researcher from assessing the influence of inattentiveness on study findings and conclusions. If researchers want to ensure a minimum number of considerate responses and the cost per response is not adjusted for attentiveness, then keeping inattentive respondents in the sample may further increase overall costs.

A strategy that may reduce the financial cost of surveys to researchers is placing attention checks throughout the questionnaire (these could be simple instructed-response items or more complex IMCs, depending on technical and time restrictions, as well as types of questions used to measure variables of interest); using these checks to measure attentiveness in combination with collection of metadata such as response time (which typically can be recorded for free); and then negotiating a lower cost per response on account of the number of seemingly inattentive respondents. Simple criteria could be used in the negotiation with the polling firm, such as only counting—for the purpose of determining whether the designated sample size has been reached—responses from individuals that complete the survey within a reasonable amount of time. What constitutes a reasonable response duration can be determined by the researcher while pilot-testing the online questionnaire or during the soft launch of the survey, by looking at the relationship between total time spent completing the questionnaire and attentiveness measured

---

based on ability to pass TQs. In implementing this recommendation researchers should make sure to ask the polling firm to record all responses, including those given by respondents completing the survey within less than the designated time minimum. Subsequently, summary information on respondent attentiveness can be incorporated into analyses of attitudes and behavior reported by the entire sample of respondents, using statistical techniques suitable for the data and research question at hand.

6 Conclusion

Using data from a recent online survey that included TQs, we evaluated the prevalence and implications of survey inattentiveness. Our results show that many respondents pay little attention to survey questions in self-completion surveys. Younger and less educated respondents, in particular, are more likely to fail TQs. Inattentives exhibit many of the symptoms of survey satisficing, including speeding and higher frequency of "don't know" responses. We also studied whether ignoring respondent attentiveness may lead to a biased evaluation of the incidence of critical attitudes and behavior. We found that when asked directly about attitudes and behavior, inattentives provide lower-intensity responses; this is also the case when they are interrogated indirectly about sensitive issues. The results of our double-list experiment suggest that inattentiveness is associated with shifts in the propensity to select sensitive items and that the presence of inattentives could challenge fundamental assumptions underlying the experimental design. On the whole, these results show that ignoring inattentive survey respondents risks significant biases in attitudinal and behavioral models.

We argue that researchers should take attentiveness seriously in survey-based studies of political behavior. In the end, what to do with inattentive survey respondents comes down to a question of survey costs relative to survey errors. Evaluating respondent attentiveness using attention checks comes at a cost. The increased questionnaire length and completion time, caused by the addition of survey items, may lead to greater respondent fatigue, administration expenses, and may influence responses to later question (Anduiza and Galais 2016). On the other hand, while preventing inattentives from completing the survey may reduce noise and bring down the cost of administering a survey, this may make subsequent analysis of these samples more complicated as they may require reweighting or other types of statistical adjustment to enable population-level inferences.

It may also be possible to learn about attentiveness by looking at survey metadata and response patterns, including the time it takes respondents to answer specific questions or to go over the entire questionnaire, as well as by examining the frequency of straightlining and tendency to report nonattitudes or unreasonable responses. More research needs to be done to assess the extent to which alternative indicators provide complimentary information about attentiveness and develop methods to combine information from numerous indicators—including different types of TQs, varying in terms of difficulty and type of challenge—into useful indicators of overall attentiveness; and to establish guidelines for incorporating this information into standard data analyses.

Supplementary materials

For supplementary materials accompanying this paper, please visit https://doi.org/10.1017/pan. 2018.57.

References

Alvarez, R. M. 1997. *Information and Elections*. Ann Arbor, MI: University of Michigan Press.
Alvarez, R. M., and J. Brehm. 2002. *Hard Choices, Easy Answers: Values, Information, and American Public Opinion*. Princeton, NJ: Princeton University Press.

Alvarez, R. M., and C. H. Franklin. 1994. "Uncertainty and Political Perceptions." *The Journal of Politics* 56(3):671–688.

Alvarez, R. M., L. R. Atkeson, I. Levin, and Y. Li. 2018. "Replication Data for: Paying Attention to Inattentive Survey Respondents." https://doi.org/10.7910/DVN/TUUYLQ, Harvard Dataverse, V1, UNF:6:ZHc1mHgkrXEorZvXXJnURQ== [fileUNF].

Anduiza, E., and C. Galais. 2016. "Answering Without Reading: IMCs and Strong Satisficing in Online Surveys." *International Journal of Public Opinion Research* 29(3):497–519.

Ansolabehere, S., and B. F. Schaffner. 2018. "Taking the Study of Political Behavior Online." In *The Oxford Handbook of Polling and Survey Methods*, edited by L. R. Atkeson and R. M. Alvarez, 76–96. New York: Oxford University Press.

Atkeson, L. R., and A. N. Adams. 2018. "Mixing Survey Modes and Its Implications." In *The Oxford Handbook of Polling and Survey Methods*, edited by L. R. Atkeson and R. M. Alvarez, 53–75. New York: Oxford University Press.

Atkeson, L. R., A. N. Adams, and R. M. Alvarez. 2014. "Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys." *Political Analysis* 22(3):304–320.

Bailey, M. A. 2017. "Selection Sensitive Survey Design: Moving Beyond Weighting." Presented at the 2017 Annual Meetings of the American Political Science Association, San Francisco, CA.

Barber, L. K., C. M. Barnes, and K. D. Carlson. 2013. "Random and Systematic Error Effects of Insomnia on Survey Behavior." *Organizational Research Methods* 16(4):616–649.

Berinsky, A. J., G. A. Huber, and G. S. Lenz. 2012. "Evaluating Online Labor Markets for Experimental Research: Amazon.Com's Mechanical Turk." *Political Analysis* 20(3):351–368.

Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3):739–753.

Berinsky, A. J., M. F. Margolis, and M. W. Sances. 2016. "Can We Turn Shirkers into Workers? *Journal of Experimental Social Psychology* 66:20–28.

Blair, G., and K. Imai. 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20(1):47–77.

Bowling, N. A. et al. 2016. "Who Cares and Who is Careless? Insufficient Effort Responding as a Reflection of Respondent Personality." *Journal of Personality and Social Psychology* 111(2):218–229.

Clifford, S., and J. Jerit. 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias? *Public Opinion Quarterly* 79(3):790–802.

Curran, P. G. 2016. "Methods for the Detection of Carelessly Invalid Responses in Survey Data." *Journal of Experimental Social Psychology* 66:4–19.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2009. *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. 3rd edn. Hoboken, NJ: John Wiley & Sons.

Downes-Le Guin, T. 2005. "Satisficing Behavior in Online Panelists." Presented at the MRA Annual Conference & Symposium, Chicago, IL.

Droitcour, J. et al. 1991. "The Item Count Technique as a Method of Indirect Questioning: A Review of its Development and a Case Study Application." In *Measurement Errors in Surveys*, edited by P. P. Biemer et al., 185–210. Hoboken, NJ: John Wiley & Sons.

Eady, G. 2016. "Replication Data for: The Statistical Analysis of Misreporting on Sensitive Survey Questions." https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PZKBUX (September 16, 2018).

Eady, G. 2017. "The Statistical Analysis of Misreporting on Sensitive Survey Questions." *Political Analysis* 25(2):241–259.

Edwards, A. L. 1957. *The Social Desirability Variable in Personality Assessment and Research*. Fort Worth, TX: Dryden Press.

Fisher, R. J. 1993. "Social Desirability Bias and the Validity of Indirect Questioning." *Journal of Consumer Research* 20(2):303–315.

Glynn, A. N. 2013. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77(S1):159–172.

Groves, R. M., and L. Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74(5):849–879.

Huang, J. L. et al. 2012. "Detecting and Deterring Insufficient Effort Responding to Surveys." *Journal of Business and Psychology* 27(1):99–114.

Imai, K. 2011. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106(494):407–416.

Johnson, J. A. 2005. "Ascertaining the Validity of Individual Protocols from Web-based Personality Inventories." *Journal of Research in Personality* 39(1):103–129.

Jones, M. S., L. A. House, and Z. Gao. 2015. "Attribute Non-Attendance and Satisficing Behavior in Online Choice Experiments." *Proceedings in Food System Dynamics* 2015:415–432.

Kapelner, A., and D. Chandler. 2010. "Preventing Satisficing in Online Surveys: A 'Kapcha' to Ensure Higher Quality Data." In *Proceedings of CrowdConf 2010, San Francisco, CA.*

Kiewiet de Jonge, C. P., and D. W. Nickerson. 2014. "Artificial Inflation or Deflation? Assessing the Item Count Technique in Comparative Surveys." *Political Behavior* 36(3):659–682.

King, G., J. Honaker, A. Joseph, and K. Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1):49–69.

Krosnick, J. A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3):213–236.

Little, R. J. A. 1992. "Regression with Missing X's: a Review." *Journal of the American Statistical Association* 87(420):1227–1237.

Maccoby, E. E., N. Maccoby, and G. Lindzey. 1954. "The Interview: A Tool of Social Science." In *Handbook of Social Psychology*, edited by G. Lindzey, 449–487. Reading, MA: Addison-Wesley.

Maniaci, M. R., and R. D. Rogge. 2014. "Caring About Carelessness: Participant Inattention and its Effects on Research." *Journal of Research in Personality* 48:61–83.

Meade, A. W., and S. B. Craig. 2012. "Identifying Careless Responses in Survey Data." *Psychological Methods* 17(3):437–455.

Miller, J. 2006. "Research Reveals Alarming Incidence of 'Undesirable' Online Panelists." Research Conference Report, RFL Communications, Skokie, IL, USA. Available at http://www.burke.com/Library/Articles/Jeff%20Miller%20RCR%20PDF.pdf.

Miller, J. D. 1984. "A New Survey Technique for Studying Deviant Behavior." PhD thesis, George Washington University.

Oppenheimer, D. M., T. Meyvis, and N. Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45(4):867–872.

Pepinsky, T. B. 2018. "A Note on Listwise Deletion Versus Multiple Imputation." *Political Analysis* 26(4):480–488.

Rosenfeld, B., K. Imai, and J. N. Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* 60(3):783–802.

Simon, H. A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63(2):129–138.

Vannette, D. 2017. "Using Attention Checks in Your Surveys May Harm Data Quality." *Qualtrics*, https://www.qualtrics.com/blog/using-attention-checks-in-your-surveys-may-harm-data-quality/ (June 14, 2018).

Ward, M. K., and S. B. Pond. 2015. "Using Virtual Presence and Survey Instructions to Minimize Careless Responding on Internet-Based Surveys." *Computers in Human Behavior* 48:554–568.

Warner, S. L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60(309):63–69.

Zagorsky, J. L., and P. Rhoton. 2008. "The Effects of Promised Monetary Incentives on Attrition in a Long-Term Panel Survey." *Public Opinion Quarterly* 72(3):502–513.