

Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?

Scott Clifford (corresponding author)

Assistant Professor

Department of Political Science

University of Houston

Philip Guthrie Hoffman Hall, 4th Floor

713-743-3890

saclifford@uh.edu

Jennifer Jerit

Associate Professor

Department of Political Science

Social and Behavioral Sciences Building, 7th Floor

Stony Brook University

Stony Brook, NY 11794

jennifer.jerit@stonybrook.edu

631-632-3267

Accepted for publication in *Public Opinion Quarterly*

SCOTT CLIFFORD is an assistant professor in the Department of Political Science at University of Houston, Houston, TX, USA. JENNIFER JERIT is an associate professor in the Department of Political Science at Stony Brook University, Stony Brook, NY, USA. The authors thank Jason Barabas, Adam Berinsky, John Bullock, Rosmery Cruz, Cindy Kam, Yanna Krupnikov, Thomas Leeper, Adam Seth Levine, Spencer Piston, Kathleen Searles, and Nick Valentino for helpful comments on earlier versions of this paper. We also thank Bob Blaemire at Catalist for assistance with the validated data.

*Address correspondence to Scott Clifford, University of Houston, Department of Political Science, Houston, TX, USA; email: saclifford@uh.edu.

Abstract

In this research note, we investigate the effectiveness of *warnings* as a method for increasing respondent motivation and decreasing survey satisficing. Four different styles of warning messages are examined in data from a randomized survey experiment conducted on the Internet. The analysis shows that three of the four warnings significantly improve respondent engagement. There is some evidence, however, that warning messages increase socially desirable responding (SDR) for certain types of people. We conclude that warnings can be a useful method for increasing attention, but that researchers should first establish that these protocol do not have unintended consequences, either for the entire sample or for theoretically relevant subgroups.

Social scientists increasingly collect data online through self-administered questionnaires. However, these data can suffer from a variety of problems—e.g., no opinion responding, non-differentiation, acquiescence—in part because respondents may not pay attention as they answer a questionnaire (Krosnick 1991; Berinsky, Margolis, and Sances 2014). This study examines the effectiveness of *warnings* as a method for improving data quality in self-administered surveys. Generally speaking, warnings encourage respondents to follow instructions or to think carefully (Krosnick 2000). There is evidence that such messages can be effective (Huang et al. 2012), but warnings may increase social desirability pressures. Moreover, this effect may be concentrated among particular respondents, such as the highly educated, because of their greater overall tendency to offer socially desirable responses (Silver, Anderson, and Abramson 1986).

Warning Messages and the Question Answering Process

Satisficing occurs when respondents do “just enough to satisfy the survey request, but no more” (Krosnick 2000, 4). Warning messages seek to reduce the likelihood of satisficing by increasing respondent motivation to generate an optimal answer to survey questions.

Warnings have been adopted in recent studies (e.g., Press, Sagan, and Valentino 2013), but these protocol can have unintended consequences. In particular, warnings may heighten respondents’ sense of surveillance, increasing concern with self-presentation. People are less likely to report socially unacceptable behaviors when a third party (e.g., the interviewer) is present (Aquilino 1992; Tourangeau and Smith 1996). Even subtle cues—such as an image of “watching eyes”— can induce reputation-management concerns which manifest in more socially desirable attitudes and behavior (e.g., Haley and Fessler 2005).

In addition, warnings may make socially desirable responding (SDR) more likely by increasing a person's level of cognitive engagement with the questionnaire. SDR is an effortful process (Tourangeau and Rasinski 1988), with individuals reaching a judgment but then "editing" their answer in response to social desirability concerns (Holtgraves 2004). Indeed, subjects who are also engaged in other tasks are *less* likely to engage in SDR (Riemer and Shavitt 2011). In this way, warnings may precipitate SDR through increased cognitive engagement. Based on the existing literature, we hypothesize that warning messages will boost respondent engagement (H1), but that they may inadvertently increase the likelihood of socially desirable responding (H2).

Some scholars view socially desirable responding as a "response strategy reflecting the sensitivity of *specific items for specific individuals*" (Tourangeau and Yan 2007, 860, emphasis added). Numerous studies have shown that the highly educated are more likely to be aware of socially acceptable behavior and feel greater pressure to appear in conformity with social norms (e.g., Karp and Brockington 2005), especially on questions relating to political participation (Silver, Anderson, and Abramson 1986; Fullerton, Dixon, and Borch 2007; Ansolabehere and Hersh 2012). Thus, we expect that the effect of warnings on socially desirable responding will be greatest for respondents with high levels of education (H3).

Experimental Design

We tested our hypotheses with a randomized experiment administered online through the Time-sharing Experiments for the Social Sciences (TESS) platform ($N = 1168$). The study was a

between-subjects design with four warning conditions and a control group.¹ In the first treatment condition (“Audit”), the message implicitly invokes the researcher:

We check responses carefully in order to make sure that people have read the instructions for the task and responded carefully. We try to only use data from participants who clearly demonstrate that they have read and understood the survey. Again, there will be some very simple questions in what follows that test whether you are reading the instructions. If you get these wrong, we may not be able to use your data. Do you understand?

The second condition (“Anonymity”) is identical to the first, but the warning includes a reminder that responses will remain anonymous (see the Appendix for complete wording). There is evidence that such assurances can boost response rates and reduce misreporting on sensitive items (e.g., Singer, VonThurn, and Miller 1995; Tourangeau and Yan 2007). Thus, the second treatment has all the satisficing-reducing properties of the original warning along with language intended to alleviate social desirability concerns (but see Lelkes et al. 2012).

The third treatment (“Commitment”) builds on research showing that people seek to be consistent with their previous commitments (Cialdini et al. 1978; Cannell, Miller, and Oskenbeurg 1981). In this condition, the stimulus reads, “It is important to us that participants in our survey pay close attention to the materials. Are you willing to carefully read the materials and answer all of the questions to the best of your ability?” and provides “Yes” or “No” responses. The fourth treatment (“Feedback”) appeals to respondents’ desire for self-insight by providing feedback on their performance at the end of the questionnaire (Gosling et al. 2004). The stimulus stated: “Many of our participants have expressed interest in the results of our

¹ Knowledge Networks (KN) conducted the survey from November 8–22, 2012. The recruitment rate (RECR) was 14.7%, the profile rate (PROR) was 65.1%, and the study-specific completion rate (COMR) was 67.7%, for a cumulative response rate of 6.1% (RECR x PROR x COMR; Callegaro and DiSogra 2008).

surveys. In order to provide you with this information, at the end of the survey you will be shown how your opinions on various topics compare to the nation as a whole.” Opt-in online surveys that motivate participation through the promise of feedback have become common in psychology, but there has been little analysis of whether such promises affect responses. The fifth condition is a control group in which respondents do not view any message. The outcome measures consist of questions designed to capture SDR and attention (Figure A1 of the Supplementary Online Materials shows a schematic of the design).

To identify SDR, we used questions previously identified as being sensitive to social desirability bias (e.g., Presser and Stinson 1998; Chen, Lee-Chai, and Bargh 2001; Holbrook, Green, and Krosnick 2003; Streb et al. 2008). The question wording is shown in the Appendix, but in brief, we included items in which there was a strong expectation of socially desirable responding (e.g., the racial resentment scale) along with questions where the effects may be more muted (e.g., texting while driving), as well as instances where SDR was expected to result in overreporting (e.g., religious attendance) and underreporting (e.g., receiving government assistance). These design features help establish the range of effects that might occur in studies utilizing warnings. A frequently-noted limitation of research on socially desirable responding is the lack of validation data (i.e., the researcher assumes that any shift toward the desirable attitude is evidence of SDR). Working with a commercial vendor (Catalist), we obtained *behavioral* data for two of our outcomes (registration status and 2010 vote). The validated data provide a benchmark against which we can assess the accuracy of the self-reported measures, allowing us to more firmly establish whether socially desirable responding is taking place (Kreuter et al. 2008).

We measure attention with two instructional manipulation checks (IMCs), located in the middle and end of the questionnaire (Oppenheimer, Meyvis, and Davidenko 2009; Berinsky, Margolis, and Sances 2014). IMCs look similar to other survey items, but the question instructs respondents to *ignore* the standard response format and to select a particular option(s). A respondent “fails” an IMC when he or she answers the question sincerely, which happens if a person has skimmed or ignored the question stem. In addition to the IMCs, we examined the frequency of item non-response and time spent on the survey as indicators of respondent engagement. Item non-response was operationalized as the number of items a person skipped. Length of time was operationalized as the log of the average for each question (excluding warnings and both IMCs), with outliers recoded to three standard deviations above the mean prior to creating the index. Insofar as warnings increase motivation, they should reduce item non-response, but increase the time spent answering questions and the passage rate on the IMCs.²

Empirical Results

Figure 1 shows the percentage of people passing the first and second IMCs across experimental conditions. Beginning with the grey bar, only 38 percent of respondents in the control group passed the first IMC. Three of the four warnings result in a statistically significant increase in passage rates, with effect sizes ranging from 8 percentage points (Commitment) to approximately 27 percentage points (Audit, Anonymity). Chi-square tests indicate that Audit and Anonymity both have larger effects than Feedback and Commitment ($p < .01$; all tests two-tailed unless indicated otherwise).

² IMC passage is associated with longer survey time and lower item non-response (both $ps < .001$).

Figure 1 here.

On the second IMC (bottom panel), a similar pattern emerges with the same three warnings resulting in a statistically significant increase in passage relative to the control condition (all $ps < .01$). Effect sizes range from 12 to 15 percentage points.

Summarizing across the indicators of attention, Audit had the most consistent effect: it increased IMC passage ($p < .01$) and time spent on the survey ($p < .05$), and decreased item non-response ($p = .05$).³ Anonymity had similar, but slightly smaller effects. That message type increased IMC passage and decreased item non-response ($p < .01$ and $p < .05$, respectively), but it had no effect on time spent on the survey. Commitment improved performance on the IMC ($p = .07$), but had null effects for time spent on the survey and item non-response. Finally, the promise of feedback had null effects on all of our indicators of attention. As we elaborate below, the Feedback warning had other unintended consequences that recommend against its use.

Three of the four warnings improve performance on IMCs and one warning in particular (Audit) increased engagement on all three measures of attention. The next series of analyses investigates whether these messages have any adverse effects. Our indicator of SDR is whether the respondent provided the socially desirable response (coded 1; 0 otherwise) on the seven items in our questionnaire (e.g., saying that one voted in the 2010 election or expressing racially liberal attitudes).

In analyses in which each outcome is analyzed separately, none of the warnings consistently increase average levels of SDR (see Table A2 in Supplemental Online Materials).

³ The results for the model predicting IMC passage and for the analyses of item non-response and time spent taking the survey are shown in the Supplementary Online Materials (Table A1).

We then created a summary measure representing the number of socially desirable answers given by a respondent. Table 1 shows the results of a regression model predicting this summary measure as a function of treatment condition.

Table 1 here.

Focusing on the first column of results (Model 1), the top portion of Table 1 shows the coefficients on the treatment indicators. Contrary to H2, there are no significant main effects for any of the warning messages on SDR.

Our third hypothesis predicts that the effects of warnings on SDR will be concentrated among the highly educated, which we test with an interaction between the treatment indicators and a four-point ordinal measure of education (1=did not finish high school; 4=bachelor's degree or higher). As expected, there are statistically significant interactions between education and the Commitment and Feedback warnings ($ps < .01$). In both cases, the warnings led highly educated respondents to report .4 additional SDR responses ($ps = .02$).⁴ Table 1 also shows that three of the four warnings *decreased* socially desirable responding among the least educated, a pattern that may indicate more accurate reporting among this subgroup. We return to this finding below, in our discussion of the behavioral data.

While the warnings may have increased socially desirable responding among the educated, an alternative explanation is that the treatment increases accuracy among all

⁴ Alternative models (e.g., repeated measures, count models) yielded similar results. For some of our questions, the socially desirable response is also the acquiescent response. We find a similar pattern of results with an item for which the acquiescent response is the *undesirable* response (Text Messaging in Table A2).

respondents, since a person's level of education is associated with holding the socially desirable attitude. The validated data will help determine whether warning messages are causing the well-educated to give more accurate, as opposed to more socially desirable, opinions. We focus on overreporting, which occurred if someone reported being registered/voting in the KN survey but was validated to be unregistered or to have not voted.

We predict overreporting on registration status and 2010 vote with a probit model. Because people who have engaged in either activity cannot overreport, we limit our attention to respondents who were unregistered or who did not vote in the 2010 general election according to administrative records ($n = 163$ and 565 , respectively). The dependent variable takes on a value of "1" if someone overreports being registered/voting, and a value of "0" if a respondent accurately reports not being registered/not having voted. In our sample, 39% of those who are not registered to vote reported being registered, while 34% of those who did not vote reported voting. Once again, we predict overreporting with the treatment indicators and the treatmentXeducation interactions. Consistent with Table 1, there is a statistically significant interaction for three of the four warnings in the registration model and a statistically significant interaction for the Feedback warning in the vote model (all $p < .05$; see Supplementary Online Materials for model estimates). The marginal effects from the model appear in Figure 2 (shown separately for respondents with a high school diploma and a college degree).

Figure 2 here.

The estimates represent the marginal effect of the warnings on the likelihood of overreporting, relative to the control condition. The error bars are large because of the small number of respondents, but the pattern is consistent with earlier results. Warning messages have a disproportionate effect on the highly educated and this pattern is most dramatic in the Feedback

condition (31-point increases in overreporting registration and voting; $p = .11$ and $p < .05$, respectively).⁵ Additionally, we find some evidence that the treatments *decreased* overreporting of registration among the less-educated. Consistent with Table 1, warnings seem to have increased accuracy among the least educated. The result was unexpected, but it may reflect increased attention among a subgroup unlikely to exhibit socially desirable responding.

Warnings have two contrasting effects: increased overreporting among the well-educated as well as decreased overreporting among the less-educated. The net result is an inflation of the correlation between education and turnout. According to the validated data, the “true” correlation between registration status and education is $.15$ ($p < .001$). Looking at the self-reported data, the analogous correlation is $.12$ in the control group and ranges from $.25$ to $.36$ in the treatment conditions. Likewise, the correlation between the validated measure of 2010 vote and education is $.22$ ($p < .001$). By contrast, in the KN data, the correlation between 2010 vote and education is $.13$ in the control group and ranges from $.17$ to $.42$ in the treatment conditions. Even if warnings do not affect mean levels of a dependent variable, they may artificially strengthen or weaken its association with key independent variables.

Conclusion

Warnings are a useful tool because they seek to improve motivation at the outset of a survey, as opposed to the post hoc exclusion of people who pay insufficient attention (e.g., Maniaci and Rogge 2014). At the same time, warning messages increase socially desirable

⁵ We observe null effects in models predicting *underreporting* at all levels of education, lending further support to the social desirability interpretation (see Supplementary Online Materials).

responding for certain people, so researchers should first establish that these protocol do not have unintended consequences for theoretically relevant subgroups.

Our analysis illustrates that researchers must also consider *how* attention is manipulated. The Audit message had the most consistent effect across the various indicators of attention. Given recent evidence that anonymity messages may decrease the accuracy of people's responses (Lelkes et al. 2012) and generally smaller effects for the Commitment warning, the Audit message appears to be most effective at increasing respondent motivation. The Feedback message did little to increase respondent engagement, and it also had some of the strongest effects on the overreporting of socially desirable behaviors. Psychologists have collected data from hundreds of thousands of people using variants of the feedback message (e.g., at websites like Project Implicit and YourMorals.org), yet it was the most reactive of the warnings we studied. The promise of feedback may be an effective way to recruit subjects, but the source of motivation (desire for self-insight) may compromise the accuracy of the measures by heightening respondents' sensitivity to their own responses.

Appendix

Question Wording

[Respondents randomly assigned into one of five conditions: an untreated control group or one of the four experimental groups shown below.]

*Treatment Condition 1: Audit*⁶

We check responses carefully in order to make sure that people have read the instructions for the task and responded carefully. **We will only accept participants who clearly demonstrate that they have read and understood the survey.** Again, there will be some very simple questions in what follows that test whether you are reading the instructions. If you get these wrong, we may not be able to use your data. Do you understand?

<1> Yes, I understand

<2> No, I do not understand

⁶ The inspiration for the Audit warning language comes from an earlier versions of the Berinsky, Margolis and Sances (2014) paper and personal communication with the lead author. Three of the treatments included a question as part of the warning. In the Audit condition, 3 people (1%) responded “don’t understand” and 1 skipped the question. In the Anonymity group, 4 people (2%) said “don’t understand” (2%) and 1 skipped. In the Commitment condition, 3 people (1%) said “No” and none refused.

Treatment Condition 2: Anonymity

We check responses carefully in order to make sure that people have read the instructions for the task and responded carefully. **Your responses are anonymous, but we will only accept participants who clearly demonstrate that they have read and understood the survey.**

Again, there will be some very simple questions in what follows that test whether you are reading the instructions. If you get these wrong, we may not be able to use your data. Do you understand?

<1> Yes, I understand

<2> No, I do not understand

Treatment Condition 3: Commitment

It is important to us that participants in our survey pay close attention to the materials. Are you willing to carefully read the materials and answer all of the questions to the best of your ability?

<1> Yes

<2> No

Treatment Condition 4: Feedback

Many of our participants have expressed interest in the results of our surveys. In order to provide you with this information, at the end of the survey you will be shown how your opinions on various topics compare to the nation as a whole.

[Subjects in Condition 4 will receive an additional message regarding their feedback at the end of the study. Their responses were compared to a national average on select items from the survey (i.e., the treatment was administered without deception).]

Condition 5: Control

No text

1. Are you currently registered to vote?

<1> Yes, I am registered to vote at my present address

<2> Yes, I am registered to vote at a different address

<3> No, I am not registered to vote

<4> I am not sure

2. Did you happen to vote in the 2010 Congressional election?

<1> Yes

<2> No

3. If your party nominated a generally well-qualified person for president who happened to be a homosexual, would you vote for that person?

<1> Yes, would

<2> No, would not

[A random half of the respondents receive question 4 before the Racial Resentment battery (question 5a-5c); the other half receive question 4 after the Racial Resentment battery.]

4. Instructional Manipulation Check #1

We are also interested in what sections people like to read in the newspaper. What people read in the paper might affect their opinions on current events. We also want to see if people are reading the questions carefully. To show that you've read this much, please mark both the "Classified" and "None of the above" boxes below. That's right, just select these two options only.

Regardless of how frequently you read the newspaper, what would you say are your favorite newspaper sections to read? Please check all that apply.

<1> National

<2> Local

<3> Classified

<4> Sports

<5> Business

<6> Science and Technology

<7> Opinion

<8> None of the above

5. To what extent do you agree or disagree with the following statement:

a. The Irish, Italians, Jews, Vietnamese and other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.

b. It's really a matter of some people not trying hard enough; if blacks would only try harder they could be just as well off as whites.

c. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class.

<1> Strongly agree

<2> Agree

<3> Slightly agree

<4> Neither agree nor disagree

<5> Slightly disagree

<6> Disagree

<7> Strongly disagree

6. And now just a few questions about you...In the past seven days, have you sent or read a text message while driving?

<1> Yes

<2> No

7. Have you or members of your immediate family ever received any of the following forms of governmental assistance? Please check all that apply.

<1> Unemployment insurance

<2> Medicare

<3> Medicaid

<4> Government subsidized housing

<5> Welfare

<6> Food stamps

<7> None of the above

8. How often do you attend religious services?

<1> More than once a week

<2> Once a week

<3> A few times a year

<4> Once a year or less

<5> Never

9. Questions sometimes have different effects on people. We'd like your opinions about the questions in this survey. How likely is it that any of the questions might make people feel as if they are being watched or monitored?

<1> Extremely likely

<2> Very likely

<3> Moderately likely

<4> Not too likely

<5> Not at all likely

10. Instructional Manipulation Check #2

Most modern theories of decision making recognize the fact that decisions do not take place in a vacuum. Individual preferences and knowledge, along with situational variables, can greatly impact the decision process. In order to facilitate our research on decision making we are interested in whether you actually take the time to read the directions. So, rather than answering

the following question accurately, please check only “Mauritania” and “None of the above” and then continue. This will demonstrate to us that you have taken the time to read and follow the instructions.

Which of the following countries have you heard mentioned in the news in the last week?

(Please check all that apply.)

<1> France

<2> Germany

<3> Switzerland

<4> Mauritania

<5> Syria

<6> Iran

<7> Canada

<8> Mexico

<9> None of the above

Supplementary Data

Supplementary data are freely available online at <http://poq.oxfordjournals.org/>.

References

- Ansolabehere, Stephen, and Eitan Hersh. 2012. "Validation: What Big Data Reveal about Survey Misreporting and the Real Electorate." *Political Analysis* 20(4): 437–59.
- Aquilino, William S. 1992. "Telephone Versus Face-to-Face Interviewing for Household Drug Use Surveys." *International Journal of the Addictions* 27(1): 71-91.
- Berinsky, Adam J., Michele Margolis, and Michael Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Internet Surveys." *American Journal of Political Science*, Forthcoming.
- Callegaro, Mario, and Charles DiSogra. 2008. "Computing Response Metrics for Online Panels." *Public Opinion Quarterly* 72(5): 1008–32.
- Cannell, Charles F., Peter V. Miller, and Lois Oskenbeurg. 1981. "Research on Interviewing Techniques." *Sociological Methodology* 12: 389–437.
- Chen, Serena, Annette Y. Lee-Chai, and John A. Bargh. 2001. "Relationship Orientation as a Moderator of the Effects of Social Power." *Journal of Personality and Social Psychology* 80(2): 173–87.
- Cialdini, Robert B., John T. Cacioppo, Rodney Bassett, and John A. Miller. 1978. "Low Ball Procedure for Producing Compliance: Commitment then Cost." *Journal of Personality and Social Psychology* 36(5): 463–76.

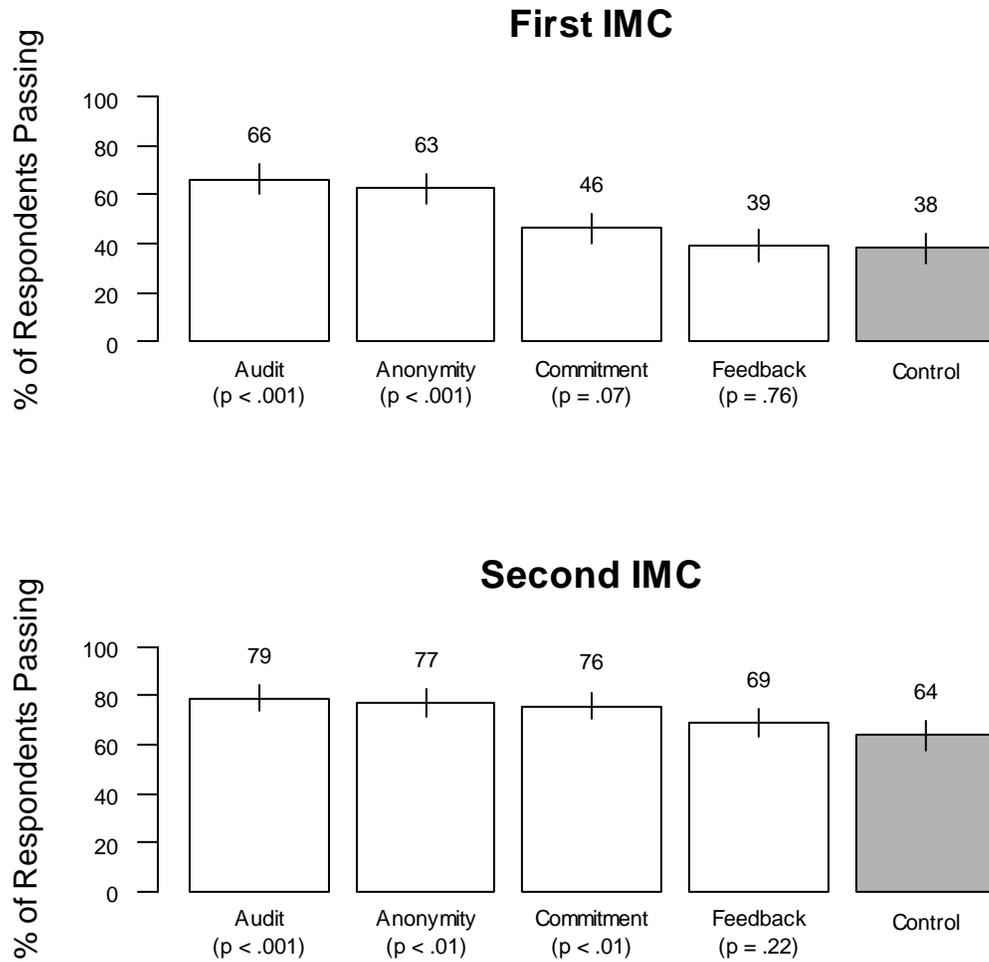
- Fullerton, Andrew W., Jeffrey C. Dixon, and Casey Borch. 2007. "Bringing Registration into Models of Vote Overreporting." *Public Opinion Quarterly* 71(4): 649–60.
- Gosling, Samuel D., Simine Vazire, Sanjay Srivastava, and Oliver P. John. 2004. "Should We Trust Web-based Studies? A Comparative Analysis of Six Preconceptions about Internet Questionnaires." *American Psychologist* 59(February/March): 93–104.
- Haley, Kevin J., and Daniel M.T. Fessler. 2005. "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Dictator Game." *Evolution and Human Behavior* 26(3): 245–56.
- Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone vs. Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Bias." *Public Opinion Quarterly* 67(1): 79–125.
- Holtgraves, Thomas. 2004. "Social Desirability and Self-Reports: Testing Models of Socially Desirable Responding." *Personality and Social Psychology Bulletin* 30(2): 161–72.
- Huang, Jason L., Paul G. Curran, Jessica Keeney, Elizabeth M. Poposki, and Richard P. DeShon. 2012. "Detecting and Deterring Insufficient Effort Responding to Surveys." *Journal of Business and Psychology* 27(1): 99–114.
- Karp, Jeffery A., and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Turnout." *Journal of Politics* 67(3): 825–40.

- Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. "Social Desirability Bias in CATI, IVR, and Web Surveys." *Public Opinion Quarterly* 72(5): 874–65.
- Krosnick, Jon A. 2000. "The Threat of Satisficing in Surveys: The Shortcuts Respondents Take in Answering Questions." *Survey Methods Newsletter* 20(1): 4–8.
- Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5(3): 213–36.
- Lelkes, Yphtach, Jon A. Krosnick, David M. Marx, Charles N. Judd, and Bernadette Park. 2012. "Complete Anonymity Compromises the Accuracy of Self-Reports." *Journal of Experimental Social Psychology* 48(6): 1291–99.
- Maniaci, Michael R., and Ronald D. Rogge. 2014. "Caring about Carelessness: Participant Inattention and Its Effects on Research." *Journal of Research in Personality* 48: 61–83.
- Oppenheimer, Daniel, Tom Meyvis, and Nicolas Davidenko. 2009. "Instructional Manipulation Checks: Detecting Satisficing to Increase Statistical Power." *Journal of Experimental Social Psychology* 45(4): 867–72.
- Press, Daryl G., Scott D. Sagan, and Benjamin A. Valentino. 2013. "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons." *American Political Science Review* 107(1): 188–206.

- Presser, Stanley, and Linda Stinson. 1998. "Data Collection Mode and Social Desirability Bias in Self-Reported Religious Attendance." *American Sociological Review* 63(1): 137–45.
- Riemer, Hila, and Sharon Shavitt. 2011. "Impression Management in Survey Responding: Easier for Collectivists or Individualists?" *Journal of Consumer Psychology* 21(2): 157–168.
- Silver, Brian D., Barbara A. Anderson, and Paul A. Abramson. 1986. "Who Overreports Voting?" *American Political Science Review* 80(2):613–24.
- Singer, Eleanor, Dawn R. Von Thurn, and Esther R. Miller. 1995. "Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature." *Public Opinion Quarterly* 59(1): 66–77.
- Streb, Matthew J., Barbara Burrell, Brian Frederick, and Michael A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72 (1): 76–89.
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Survey Questions." *Psychological Bulletin* 133(5): 859–883.
- Tourangeau Roger, and Tom W. Smith. 1996. "Asking Sensitive Questions: The Impact of Data Collection Mode, Question Format, and Question Context." *Public Opinion Quarterly* 60(2): 275–304.

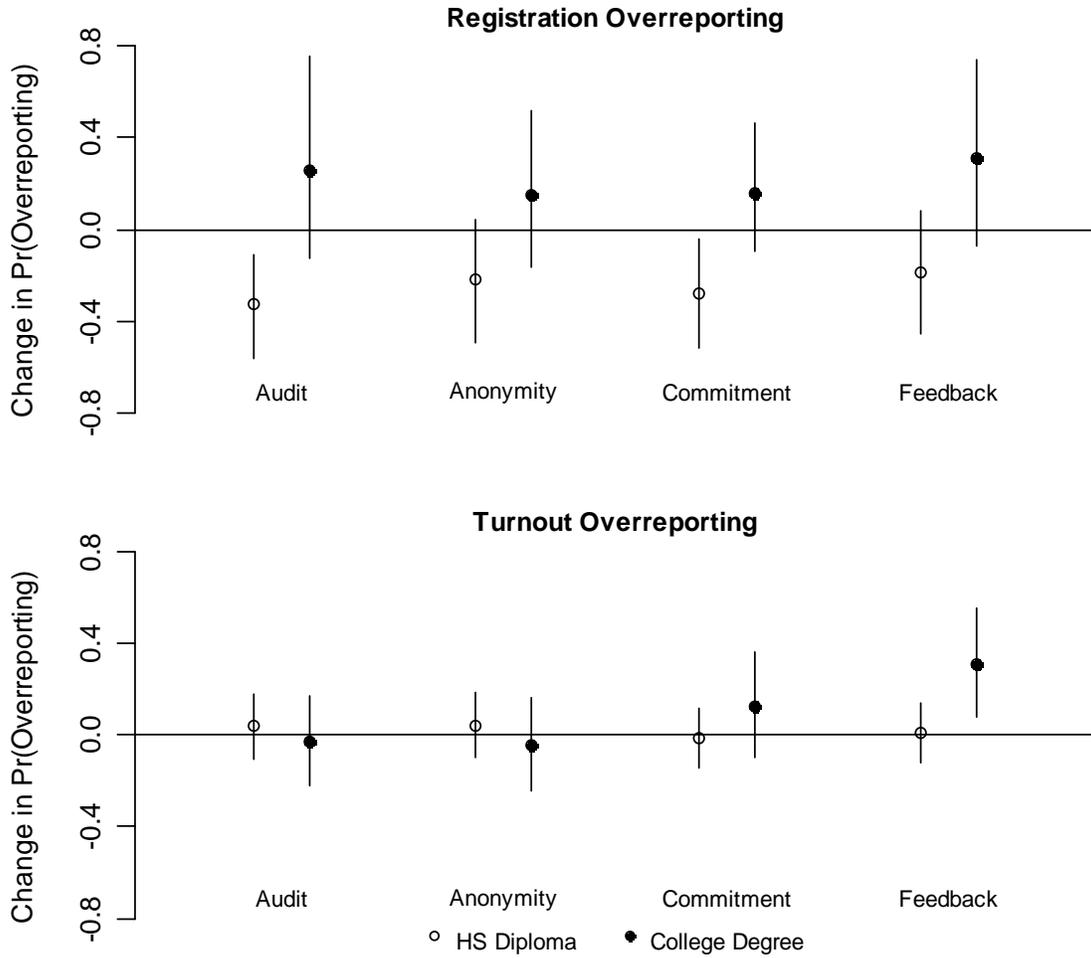
Tourangeau, Roger, and Kenneth A. Rasinski. 1988. "Cognitive Processes Underlying Context Effects in Attitude Measurement." *Psychological Bulletin* 103(3): 299–314.

Figure 1. Performance on Instructional Manipulation Checks



Notes: Figure shows percentage of respondents "passing" each instructional manipulation check by experimental condition. Reported p-values (two-tailed) come from a chi-square test comparing each treatment condition to the control.

Figure 2. Analysis of Overreporting on Vote and Registration



Notes: Estimates from models predicting overreporting on vote and registration questions among respondents who did not vote/were not registered to vote. The figure displays the estimated marginal effect of each warning condition relative to the control condition. Open circles represent effects for those earning only a high school diploma, and filled circles represent effects for those earning a college degree. Bars represent 95% confidence intervals. See text and Appendix for model details.

Table 1. OLS Model Predicting Number of Socially Desirable Responses

	Model 1 (n = 1125)	Model 2 (n = 1125)
<i>Warning Conditions</i>		
Audit	-0.15 (0.12)	-0.60 (0.34)
Anonymity	-0.01 (0.12)	-0.12 (0.35)
Commitment	0.02 (0.12)	-0.93 (0.35) **
Feedback	-0.05 (0.12)	-1.08 (0.34) **
<i>Education and Interaction Terms</i>		
Education		0.25 (0.08) **
Audit*Education		0.17 (0.12)
Anonymity*Education		0.04 (0.12)
Commitment*Education		0.34 (0.12) **
Feedback*Education		0.37 (0.11) **
Constant	4.46 (0.09) **	3.75 (0.25) **
R ²	0.00	0.13

Notes: Dependent variable is the number of socially desirable attitudes reported across the seven outcome measures. Education ranges from 1 (did not finish high school) to 4 (bachelor's degree or higher).

* $p < .05$ ** $p < .01$ two-tailed. Standard errors in parentheses.