Journal of Experimental Political Science (2020), **7**, 209–217
doi:10.1017/XPS.2019.25

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Should We Worry About Sponsorship-Induced Bias in Online Political Science Surveys?

Thomas J. Leeper* and Emily A. Thorson†

*Department of Methodology, London School of Economics, London WC2A 2AE, UK, e-mail: thosjleeper@gmail.com, **Twitter:** @thosjleeper
†Department of Political Science, Syracuse University, Syracuse, NY, USA, e-mail: ethorson@gmail.com, **Twitter:** @emilythorson

Abstract

Political scientists rely heavily on survey research to gain insights into public attitudes and behaviors. Over the past decade, survey data collection has moved away from personal face-to-face and telephone interviewing towards a model of computer-assisted self-interviewing. A hallmark of many online surveys is the prominent display of the survey's sponsor, most often an academic institution, in the initial consent form and/or on the survey website itself. It is an open question whether these displays of academic survey sponsorship could increase total survey error. We measure the extent to which sponsorship (by a university or marketing firm) affects data quality, including satisficing behavior, demand characteristics, and socially desirable responding. In addition, we examine whether sponsor effects vary depending on the participant's experience with online surveys. Overall, we find no evidence that response quality is affected by survey sponsor or by past survey experience.

Keywords: experiment; survey; bias; methodology; satisficing; social desirability

Political scientists, particularly those who use experimental methods, increasingly conduct research using online surveys. This increased reliance on a comparatively new survey format raises important questions about recruitment, respondent engagement, and data quality. A striking feature of many web-based interviews is the prominence of the study's sponsor (often an academic institution), at the start of the interview and often throughout the survey. This prominence is unique to the online format. While telephone polling has historically included a brief introductory message describing the sponsor, and face-to-face interviews involve an introductory letter or mention of the research team leading the project, online surveys (and online experiments) typically contain detailed introductory text describing the

Citation of data: Authorship is equal and listed alphabetically. This paper was previously presented at the 2015 Annual Meeting of the Midwest Political Science Association, Chicago, IL. Thanks to Brad Jones for helpful feedback. The data, code, and any additional materials required to replicate all analyses in this article are available at the *Journal of Experimental Political Science Dataverse* within the Harvard Dataverse Network, at doi:10.7910/DVN/KKFS8Y.

study, its sponsorship, the name of the principal investigator, and human subject regulations. In many surveys, the name and logo of the sponsoring academic institution appears in the header of every page of the survey questionnaire. Participants in online studies are, therefore, shown a much stronger exposure to the sponsor than in other survey modes historically used by political scientists.

While previous research has examined the effect of sponsorship, including academic sponsorship, on response rates (Tourangeau, Presser, and Sun 2014), no extant research examines how sponsorship affects response behavior in the online context. If sponsorship effects exist, they could seriously undermine the external validity of findings. Studying sponsorship effects is difficult, however, because participants may be more or less willing to participate in studies from different sponsors, thereby confounding selection effects and response biases. To test whether prominent and repeated reminders of academic research sponsorship change the behavior of online participants, we conducted an online experiment that randomly assigns sponsorship *after* participants have agreed to respond. Using a variety of measures of data quality, we find minimal influence of sponsorship by either a university or marketing firm, nor any effect of past survey experience on survey data quality. The results give us confidence in the quality of web-based data collection for political science research.

SPONSORSHIP AND BRANDING IN ONLINE RESEARCH

Relatively recently, conducting a survey required contracting with a survey firm that maintained control over many aspects of the instrument. Today, the growth of online survey tools has "democratized" social and political research, allowing investigators to program and distribute their own surveys on platforms such as Mechanical Turk, Qualtrics, Prolific Academic, social media sites, or email listservs, or by contracting with online panels to recruit respondents.

There are three main points from which respondents may learn about the sponsor of an online survey: in the recruitment message, the consent form, or the body of the survey itself. When recruiting subjects (e.g., from Mechanical Turk), many researchers choose to include their university affiliation in the recruitment text. After opting in, respondents are almost always given at least a brief exposure to the university's name in the text of the consent form as required by institutional review boards. Finally, many researcher-designed online surveys feature the university's logo at the top of each page (several examples of this practice can be found in Supplementary Material: Appendix A). In contrast to this relatively prominent display of sponsorship in online surveys, most telephone interviews name the sponsor only once, at the beginning of the interview (see Supplementary Material: Appendix A).

Past research suggests that surveys sponsored by academic institutions yield higher response rates compared with those administered by commercial firms (Fox, Crask, and Kim 1988; Jones and Linda 1978). Universities are perceived as being more trustworthy, likeable, and having higher authority status compared with commercial companies, all of which increases the likelihood that a potential respondent will accede to the request (Groves, Cialdini, and Couper 1992). Indeed, surveys

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

sponsored by within-state universities garner higher response rates compared with those sponsored by out-of-state universities (Edwards, Dillman, and Smyth 2014). However, the same factors that improve response rates for university-sponsored surveys may also bias results in several ways.

When the survey sponsor is an institution that the respondent likes and respects (such as a university), respondents may be more concerned with social desirability. "Social desirability bias" occurs when respondents misrepresent themselves in order to appear compliant with social norms (Nederhof 2006). For example, social desirability often causes people to over-report voting (Belli et al. 1996). While the size of the over-reporting varies by question wording, estimates suggest that it can run as high as 20% (McDonald 2003). In health-related surveys, social desirability can lead people to under-report behaviors such as overeating or drug use. Social desirability bias is reduced in self-administered surveys compared with telephone or in-person surveys (Kreuter, Preser, and Tourangeau 2008). This mode effect suggests that a respondent's sense of who is "listening" to their answers shapes how they respond. Prominent displays of university sponsorship could thus heighten social desirability effects.

Respondents' attitudes towards the survey sponsor might also increase demand characteristics, specifically the "good-subject effect," in which subjects behave in a way that will "help" the researcher (Weber and Cook 1972). For example, when unobtrusively told of an experiment's hypothesis, respondents were significantly more likely to behave in a way that confirmed that hypothesis, and this effect was heightened among respondents who had a positive attitude towards the experimenter (Nichols and Maner 2008).

Finally, university sponsorship might also increase attentiveness and reduce satisficing. Satisficing occurs when respondents expend minimal energy in answering questions. This may include careless reading of response options or a less effortful memory search (Krosnick 1999).

We hypothesize that a survey's sponsor affects respondents' response behavior in several ways. In particular, we expect that, compared with a no-sponsor control group, the university sponsor will increase socially desirable responding and demand effects (Nichols and Maner 2008) and decrease satisficing behavior (Krosnick 1991). Similarly, we expect that, compared with the control condition, a commercial marketing sponsor will decrease socially desirable responding and demand effects and increase satisficing.

Design

There are two challenges to assessing sponsorship effects in surveys. First, the questionnaires must be identical apart from the sponsor so that response behavior is not affected by question or response wording. Second, sponsorship may affect who is willing to participate in a survey: different sponsors may yield samples that differ in unobserved ways (due to the sponsor's reputation or variations in recruitment techniques). These differences may in turn affect response behavior. Thus, a test of sponsorship bias requires a randomized experiment where all respondents complete the same questionnaire only *after* opting in to survey participation.

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

We recruited 852 respondents from Amazon's Mechanical Turk (MTurk), an online opt-in crowdsourcing platform based in the United States.[1] While MTurk does not demographically reflect the U.S. population as a whole, it does offer two advantages. First, it is a population frequently used by social scientists to field experiments and surveys. Second, MTurk respondents (or "workers") have a wide range of previous experience: while some have taken hundreds of surveys, others have taken only a few. This variation allows us to examine whether response bias varies depending on the respondent's survey experience.

U.S.-based respondents were offered $1.00 as compensation to "complete a 10–15-minute survey about your attitudes and opinions." The MTurk "requester" (sponsor name that was displayed to potential participants) was "Aarhus Research,"[2] and no other information about the study was available to respondents until they agreed to participate in the study.

We additionally employed quota sampling to obtain a sample of respondents stratified by the amount of their experience on MTurk.[3] Each task an MTurk "worker" completes is called a HIT, and we used MTurk's "Qualification Requirements" to create separate survey opportunities (i.e., separate HITs) for workers with varying numbers of completed HITs (<100, 100–500, 500–1,000, 1,000–2,000, and >2,000).[4] We are, therefore, able to detect whether any observed effects of sponsorship are moderated by previous task experience, thereby addressing concerns about different response patterns among individuals who participate in many online surveys (Binswanger, Schunk, and Toepoel 2013).

After agreeing to participate in the survey, respondents were randomly assigned to one of five conditions. In the first (control) condition, the survey was not attributed to any particular sponsor and the survey included no logos or images. In the next two conditions, respondents were told the survey was being conducted by a marketing research firm called "Aarhus Market Research." In one of these conditions (*marketing light*), participants were briefly told the name of the sponsor and then continued with the survey. In the other (*marketing heavy*), participants were also given a longer description of the firm and asked whether they had previously completed any surveys for this firm.[5] In both heavy and light conditions, respondents saw the logo at the top of each page of the survey.

In the final two conditions, respondents were told that the survey was being conducted by the University of Aarhus and shown a university logo. In one condition (*university light*), respondents were told only the name of the university sponsor, while in the other (*university heavy*), respondents were given a longer

[1]Data available on Harvard Dataverse (Leeper and Thorson 2019).

[2]This was a new requester account created specifically for this study, so as to mitigate any reputational concerns that might affect who would participate in the study.

[3]See Supplementary Material: Appendix G for details.

[4]We aimed to recruit 150 workers at each level and an additional 150 in the >2,000 category given the lack of an upper bound for the number of completed HITs. The final size of each stratum was 150, 164, 157, 106, and 295, respectively.

[5]While this manipulation involves deception, only participants in the marketing conditions were misled about the survey sponsor, and these participants were debriefed about the manipulation after completing the study. Given that participants already agreed to participate in the survey (not knowing the sponsor), we felt that this manipulation involved bare minimum risk of harm to participants.

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

description of the university and asked whether they had participated in any research for the university before. The light and heavy versions of our two sponsor treatments were designed to test the conditionality of any effects and to better correspond to real-world survey conditions, in which the extent of exposure to the survey sponsor can vary a great deal.

Data Quality Measures

To measure the influence of sponsorship, we employed four categories of survey and experimental measures: questions likely to evoke socially desirable responding, measures of the "good subject effect," attentiveness measures, and two knowledge batteries that also serve as measures of satisficing. These sets of measures were presented to respondents in random order. See Supplementary Material: Appendix F for full question wordings.

We implemented four tests of socially desirable responding: self-reported vote history, a "double list experiment" modified from Glynn (2013), a "check all that apply" question listing a variety of socially desirable and undesirable items, and self-reported interest in politics and public affairs. The survey also included an experimental test of the "good subject effect," in which respondents alter their behavior to confirm the researcher's hypothesis after learning the purpose of a study. In this measure, adopted from Nichols and Maner (2008), respondents viewed a sequence of 10 pairs of neutral images and were asked to select the image they preferred from each pair. Half of respondents were randomly assigned to receive introductory text telling them that "we believe that when people are choosing between two images that are very much alike, they prefer images on the left." We then compared the number of "left" images selected in this group to the number of "left" images selected in a control group that did not receive the hypothesis explanation.

The third category of measures assessed attentiveness. Respondents read a short excerpt from a news article about politics and were then asked to write down everything they could remember about the text in an open-ended text box. This task yielded three variables: (1) total time reading, (2) total number of characters typed, and (3) total number of correct and incorrect pieces of information recalled.

The final category, assessing satisficing, consisted of 10 knowledge questions (both open-ended and true–false) concerning politics and scientific understanding. Every question included an explicit "don't know" option. These questions yielded two measures: the number of correct responses and the number of "don't know" responses. The advantage of using knowledge questions to assess satisficing is that they require a more effortful information search, thereby increasing the temptation to satisfice by selecting the "don't know" option. Finally, we measured effortful responding by asking respondents explicitly if they had used outside help to answer any of the knowledge questions (by using the internet to look up correct answers; see Jensen and Thomsen 2013).

The survey also included basic demographics (sex, race, education, party identification, ideology, and measures of the need for cognition and need to evaluate) and four final questions measuring perceptions of the survey interview. These included: (1) a feeling thermometer measures of respondents' attitudes towards different types of organizations, including universities and market research companies;

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

Table 1
Summary of Effects

Measure	Test for group differences
Reported past voting	F (4, 709) = 3.09, $p \leq 0.02$
Reported good behaviors	F (4, 811) = 0.09, $p \leq 0.99$
Political interest	F (4, 808) = 0.75, $p \leq 0.56$
Good-subjects behavior	F (4, 808) = 2.21, $p \leq 0.07$
Information recall (characters)	F (4, 811) = 0.78, $p \leq 0.54$
Information recall (incorrect)	F (4, 799) = 1.26, $p \leq 0.28$
Information recall (timing)	F (4, 811) = 1.18, $p \leq 0.32$
Attention check	F (4, 804) = 1.70, $p \leq 0.15$
Political knowledge (total)	F (4, 532) = 0.77, $p \leq 0.55$
Political knowledge (DKs)	F (4, 811) = 0.62, $p \leq 0.65$
General knowledge (total)	F (4, 811) = 0.62, $p \leq 0.65$
General knowledge (DKs)	F (4, 811) = 0.62, $p \leq 0.65$
Cheating: political knowledge	F (4, 810) = 0.47, $p \leq 0.76$
Cheating: general knowledge	F (4, 810) = 2.74, $p \leq 0.03$
Willing to receive email	F (4, 808) = 1.01, $p \leq 0.40$
Survey rating	F (4, 805) = 1.31, $p \leq 0.26$
Fair compensation	F (4, 811) = 0.94, $p \leq 0.44$
Evaluation of universities	F (4, 720) = 0.08, $p \leq 0.99$
Evaluation of marketing firms	F (4, 784) = 2.21, $p \leq 0.07$

Cell entries are *F*-tests for joint hypothesis of differences across experimental conditions from an OLS regression controlling for respondent experience on MTurk. Full results are included in Supplementary Material: Appendix D. DK = don't know answers.

(2) a measure assessing respondents' perceptions of the quality of the survey compared with others they have responded to; (3) an open-ended measure asking participants how much they felt they should have been paid for participation in the survey; and (4) a yes–no question asking whether respondents would be willing to receive further emails from the investigators. Participants in the marketing conditions were then debriefed.

RESULTS

Demographic characteristics of the 852 participants are available in Supplementary Material: Appendix C. To assess the impact of sponsorship on our measures of survey response behavior, we employed a series of regressions. Table 1 summarizes the results, which are available in full in Supplementary Material: Appendix D. For each

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

analysis, one measure of survey responding was regressed on indicators for each of our treatment conditions (with the control group as the baseline), controlling for respondent experience (our blocking factor). We present a joint significance test for group differences across experimental conditions, with the resulting *F*-statistic and *p*-value reported for each survey measure. Overall, we find almost no group differences, strongly suggesting that the survey sponsor has a minimal effect on participants' behavior. The effect of sponsorship did not vary by the experience level of the respondent, by the identity of the sponsor (marketing vs. university), or by the intensity of the treatment (exposure to sponsor identity only at the beginning of the survey vs. throughout).

Respondents in the marketing sponsorship conditions reported lower rates of voting compared with those in the control group or in the university sponsorship conditions. This is the only significant effect for any of the three social desirability measures. Regarding the good-subjects effect, we see a large and positive treatment effect that replicates the results of Nichols and Maner (2008): when told about the researchers' hypothesis, individuals engage in behavior that strongly conforms with that hypothesis. While this effect holds across all conditions, its magnitude is unaffected by the survey sponsor.

The knowledge and attentiveness measures show that there is no statistically significant effect of sponsorship on respondent engagement. In addition, sponsorship did not affect the rate at which respondents passed the more explicit attention-check questions. Respondents with the most MTurk experience (>2,000 HITs) were slightly less likely to give "don't know" responses to, or report cheating on, political knowledge questions.[6]

Across all five aspects of respondents' evaluations of survey experience, sponsorship has no effect. However, more experienced respondents are likely to agree to receiving emails in the future. It is also worth noting that in each of the conditions (including the control condition), respondents were significantly more favorable towards "colleges and universities" than "market research firms," which is in line with previous research suggesting that universities are perceived more positively compared with other common survey sponsors (Fox, Crask, and Kim 1988).

DISCUSSION

The results indicate that sponsorship had little effect on socially desirable responding, attentiveness, satisficing, or demand effects. These minimal effects are encouraging given the increasing use of researcher-designed online surveys in political science, although more research is needed to better understand the circumstances under which sponsorship could affect survey error. For example, interactions between the sponsor and the survey topic (such as a university-sponsored survey about student debt) could create more bias compared with the relatively neutral content in this study. Finally, we caution that while our university and market research firms did not affect response quality, effects may exist for other sponsors, especially those perceived as being partisan or having a distinct agenda (Presser, Blair, and Triplett 1992;

[6]However, very low numbers of respondents reported cheating behavior (32 for political knowledge items and 13 for science items).

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

Tourangeau et al. 2009; Tourangeau, Presser, and Sun 2014). These results are positive in that they suggest that adhering to the AAPOR (American Association of Public Opinion Research) ethical standards, which prohibit making "false or misleading claims as to a study's sponsorship or purpose," does not compromise survey quality.[7]

Several limitations should be highlighted. First, the sample here leaned towards liberal Democratic and had higher levels of education compared with the U.S. general population. It would be worth examining whether those with low levels of education might respond differently to survey sponsorship. Second, it is possible that the lack of observed effect was due to a weak treatment rather than a true lack of sponsor-induced bias. However, the "treatment" in this survey (the sponsor mentioned in the consent form as well as via a header throughout the survey) is a relatively realistic portrayal of sponsorship in an online context. Thus, while a more prominent display of the sponsor (e.g., a video showing a professor giving instructions) might, in fact, increase bias, such displays are uncommon in most real-world surveys. A similar concern might be raised about the identity of the sponsor ("Aarhus University"), which is likely unfamiliar to most respondents, weakening the treatment. However, because past research suggests that within the academic context, perceived authority/prestige does not affect response rates, we would expect similar results for other university sponsors (Porter and Whitcomb 2003).

A final limitation is the sample itself. Participants recruited via MTurk are more attentive compared to other online survey participants (Hauser and Schwarz 2016). This might affect the results in one of two ways: either by heightening the effect of the treatment (because of participants' increased attentiveness to the survey sponsor) or by reducing the magnitude of effects, because subjects are simply less likely to engage in problematic behavior. While the inclusion of respondents' level of MTurk experience mitigates these concerns by testing for differential effects by level of "professionalization," it is still possible that samples that are overall less attentive might respond differently to sponsorship. Overall, these results are encouraging for researchers aiming to reduce potential sources of error when conducting online surveys.

Supplementary Material. To view supplementary material for this article, please visit https://doi.org/10.1017/XPS.2019.25.

References

Belli, Robert F., Michael W. Traugott, Margaret Young and Katherine McGonagle. 1996. Reducing Vote Overreporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring. *Public Opinion Quarterly* *63*(1): 90–108.

Binswanger, Johannes, Daniel Schunk and Vera Toepoel. 2013. Panel Conditioning in Difficult Attitudinal Questions. *Public Opinion Quarterly* *77*(3): 783–97.

Edwards, Michelle, Don Dillman and Jolene Smyth. 2014. An Experimental Test of the Effects of Survey Sponsorship on Internet and Mail Survey Response. *Public Opinion Quarterly* *78*(3): 734–50.

Fox, Richard, Melvin Crask and Jonghoon Kim. 1988. Mail Survey Response Rate: A Meta-Analysis of Selected Techniques for Inducing Response. *Public Opinion Quarterly* *52*(4): 467–91.

Glynn, Adam N. 2013. What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment. *Public Opinion Quarterly* *77*(S1): 159–72. URL: http://poq.oxfordjournals.org/cgi/doi/10.1093/poq/nfs070

[7]AAPOR Code of Ethics, November 2015.

https://doi.org/10.1017/XPS.2019.25 Published online by Cambridge University Press

Groves, Robert, Robert Cialdini and Mick Couper. 1992. Understanding the Decision to Participate in a Survey. *Public Opinion Quarterly 56*(4): 475–95.

Hauser, David J. and Norbert Schwarz. 2016. Attentive Turkers: MTurk Participants Perform Better on Online Attention Checks than Do Subject Pool Participants. *Behavior Research Methods 48*(1): 400–407.

Jensen, Carsten and Jens Peter Frølund Thomsen. 2013. Self-Reported Cheating in Web Surveys on Political Knowledge. *Quality & Quantity 48*(6): 3343–54.URL: http://link.springer.com/10.1007/s11135-013-9960-z

Jones, Wesley H. and Gerald Linda. 1978. Multiple Criteria Effects in a Mail Survey Experiment. *Journal of Marketing Research 15*(2): 280–4.

Kreuter, Frauke, Stanley Preser and Roger Tourangeau. 2008. Social Desirability in CATI, IVR, and Web Surveys. *Public Opinion Quarterly 72*(5): 847–65.

Krosnick, Jon A. 1991. Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology 5*: 213–36.

Krosnick, Jon A. 1999. Survey Research. *Annual Review of Psychology 50*: 537–67.

Leeper, Thomas and Emily Thorson. 2019. Replication Data for: Should We Worry About Sponsorship-Induced Bias in Online Political Science Surveys? *Harvard Dataverse*. doi:10.7910/DVN/KKFS8Y.

McDonald, Michael P. 2003. On the Over-Report Bias of the National Election Survey. *Political Analysis 11*: 180–6.

Nederhof, Anton J. 2006. Methods of Coping with Social Desirability Bias: A Review. *European Journal of Social Psychology 15*(3): 263–80.

Nichols, Austin Lee and Jon K. Maner. 2008. The Good-Subject Effect: Investigating Participant Demand Characteristics. *The Journal of General Psychology 135*(2): 151–66.

Porter, Stephen R and Michael E. Whitcomb. 2003. The Impact of Contact Type on Web Survey Response Rates. *The Public Opinion Quarterly 67*(4): 579–88.

Presser, Stanley, Johnny Blair and Timothy Triplett. 1992. Survey Sponsorship, Response Rates, and Response Effects. *Social Science Quarterly 73*: 699–702.

Tourangeau, Roger, Robert Groves, Courtney Kennedy and Ting Yan. 2009. The Presentation of a Web Survey, Nonresponse and Measurement Error among Members of Web Panel. *Journal of Official Statistics 25*: 299–321.

Tourangeau, Roger, Stanley Presser and Hanyu Sun. 2014. The Impact of Partisan Sponsorship on Political Surveys. *Public Opinion Quarterly 78*(2): 510–22. URL: http://dx.doi.org/10.1093/poq/nfu020

Weber, Stephen J. and Thomas D. Cook. 1972. Subject Effects in Laboratory Research: An Examination of Subject Roles, Demand Characteristics, and Valid Infererence. *Psychological Bulletin 77*(4): 273–95.

Cite this article: Leeper TJ and Thorson EA (2020). Should We Worry About Sponsorship-Induced Bias in Online Political Science Surveys?. *Journal of Experimental Political Science* **7**, 209–217. https://doi.org/10.1017/XPS.2019.25