

SPONSORSHIP EFFECTS IN ONLINE SURVEYS

Charles Crabtree[†]
Holger L. Kern[‡]
Matthew T. Pietryka[§]

August 6, 2018

We thank Hans Hassell and Lauren Ratliff Santoro and audiences at Florida State University and the 2018 Midwest Political Science Association meeting for useful feedback. All errors remain our own.

[†]PhD cand., Department of Political Science, University of Michigan.

[‡]Assistant Professor, Department of Political Science, Florida State University.

[§]Assistant Professor, Department of Political Science, Florida State University.

ABSTRACT

Many academic surveys administered online include a banner along the top of the survey displaying the name or logo of the researcher's university. Our study aims to determine whether these banners influence survey respondents' answers, that is, whether they induce sponsorship effects. For this purpose, we field three different studies on Amazon's MTurk where we randomly assign the sponsoring institution. Our outcome measures include survey questions about social conservatism, religious practices, group affect, and political knowledge. We find that respondents provide similar answers and exhibit similar levels of effort regardless of the apparent sponsor.

Survey researchers are well aware that respondent answers are sensitive to subtle features of the survey instrument and the environment in which it is administered. Indeed, a large literature suggests that survey responses can change with minor differences in stimuli such as interviewer characteristics (Cotter, Cohen and Coulter 1982), question wording (Kinder and Sanders 1990), question order (Sigelman 1981), response-option order (Galesic et al. 2008), and even the activities the respondents were engaged in before participating in the survey (Druckman and Leeper 2012). With the rising popularity of online surveys, scholars have rushed to evaluate the validity of samples recruited online (e.g., Clifford, Jewell and Waggoner 2015; Krupnikov and Levine 2014), but less attention has been paid to the effects of the stimuli unique to online survey environments.

Online, self-administered surveys eliminate some of the potential environmentally-induced biases present in other survey modes, but they are *not* stimulus-free environments. Just as telephone interviewers provide cues about their backgrounds, attitudes, and expectations (Cotter, Cohen and Coulter 1982), online surveys provide cues about the investigators that may influence the survey response. One such cue is the sponsoring university’s name or logo, often displayed prominently in a banner at the top of the survey screen. Despite the prominence of these logos and the frequency of their use, little research has examined whether they influence responses. If they do, this influence would call into question a large body of published research because university names or logos are commonly displayed by default in the banners of surveys administered over popular survey platforms such as Qualtrics.¹

Existing research provides little clarity about the extent to which these banners might influence responses. On one hand, many scholars emphasize the instability of survey responses in the face of even seemingly imperceptible stimuli. On the other hand, several recent studies have found no discernible effect when altering small parts of the online survey experience (Mummolo and Peterson 2017; White et al. 2018). Burdein, Lodge and Taber (2006, 360) synthesize the first view, explaining that “[b]uried beneath a citizen’s answer to a survey question [...] lie myriad automatic responses

¹We contacted Qualtrics to determine how commonly universities placed their name or logo in their default survey header. Andrew Camp, a Qualtrics Product Specialist, responded through email on 2018-03-14. He confirmed that universities typically work with Qualtrics to create custom, branded survey formats, or themes, but licensing terms prevented him from disclosing descriptive statistics. He did attest that “universities do tend to prefer to set these custom themes as a default to encourage their users to use said theme.”

to environmental primes, many never consciously perceived.” And Connors, Krupnikov and Ryan (2017) provide an example of how a small detail in an online survey can affect responses. They demonstrate that disclosing in the instructions that survey responses will be shared in an *anonymized* public dataset can encourage some respondents to provide more ‘don’t know’ responses and socially desirable answers. More directly, extant work demonstrates that the purported survey sponsor can affect response rates and alter a sample’s representativeness (Edwards, Dillman and Smyth 2014; Tourangeau et al. 2009; but see White et al. 2018). By randomizing sponsorship at recruitment, however, this work cannot identify the causal effects of sponsorship on survey responses because of the potential differences between the self-selected groups. Leeper and Thorson (2015) address this point by revealing purported sponsorship only after respondents were recruited through MTurk, randomly assigning an Aarhus University logo, a marketing firm logo, or no logo at all. They find only small differences in effort and socially desirable responses between the randomly assigned groups. Of course, one might expect larger effects for universities with more salient reputations among US-based respondents, such as those typically recruited through MTurk.

To examine sponsorship effects, we fielded three studies on MTurk that varied the presence of university logos. In each study, respondents were recruited to participate before they were assigned to one of three groups. The first group saw no banner while the other two groups saw banners with one of two university logos. Study 1 used University of Notre Dame and Ohio State University logos and focused on questions about social conservatism and religiosity. Study 2 used UC Berkeley and Liberty University logos and focused on affect toward various social groups. Study 3 used Harvard University and Fitchburg State University logos and focused on political knowledge and effort. Across all three studies, we find no evidence that banners systematically influence survey responses.

SURVEY EXPERIMENTS

In each of the three studies, the experimental manipulation consisted of the prominent display of one of several banners on survey webpages. Randomization was at the respondent level, with each banner having an equal probability of being assigned. Each survey question was displayed on its own page; the same banner was shown on each of these pages for a given respondent.² Banners were sized

²In many cases, researchers display multiple questions per page. By only displaying a single question on each page we maximize the chance that participants notice the banner while responding.

to be equally tall and arranged so that they covered approximately the top third of the respondent’s screen (for examples, see SI figures A1–A3). All three studies were restricted to MTurkers with U.S.-based accounts; respondents whose IP addresses revealed that they participated from outside the 50 U.S. states were dropped from the sample. Question order and the ordering of response categories within questions were randomly assigned. The design and analysis plans for the three studies were pre-registered with the *Open Science Foundation*.³ Respondents were debriefed once all responses in a given study had been collected. Participants in earlier studies were barred from participating in later studies. Participants were paid \$0.35.

STUDY 1

We began with a difficult test, examining sponsorship effects in relation to questions on which respondents often hold strong attitudes. Study 1 asked two questions that tap into social conservatism and two questions about religious practices. Using questions from the 2012 American National Election Study and the 2016 General Social Survey, respectively, we asked about views on abortion and birth control. Using question wording from the 2016 General Social Survey, we asked about the frequency of religious attendance and prayer.⁴

Our target sample size was about 2,000 respondents. We dropped respondents whose latitude and longitude coordinates, as reported by Qualtrics, were not located within the 50 U.S. states. We also dropped respondents who did not agree to participate in the survey or who failed to complete the survey. Finally, we ordered submitted surveys by their submission times and dropped any surveys with latitude/longitude coordinates identical to previously submitted surveys. We did this so as to prevent the same individual from participating in our survey several times using different MTurk IDs. See Table A1 for details. Our realized sample size was $n = 1,646$.

We randomly assigned one of three banners to each respondent: a banner that was blank, a banner that displayed the *Ohio State University* logo, and a banner that displayed the *University of Notre Dame* logo. Notre Dame’s prominence as a Catholic university lead us to expect that seeing

This stronger-than-usual treatment implies that our inferences about the absence of sponsorship effects are likely to be conservative, making it less likely that sponsorship effects exist “in the wild.”

³Anonymous versions of our pre-registration documents can be accessed here (Study 1), here (Study 2), and here (Study 3).

⁴See Tables A2–A5 in the SI for the question wording and response options.

the Notre Dame logo might induce respondents to report more frequent religious behaviors or give more conservative answers than seeing the Ohio State logo or no logo. We included the pure control (no logo) condition to preserve our ability to distinguish between effects stemming from seeing any logo and effects stemming from seeing a specific university logo.⁵

We rely on randomization inference (RI) throughout to test the *sharp null hypothesis* of absolutely no treatment effect for any respondent (Fisher 1935; Imbens and Rubin 2015). Given the nominal level of our survey data we use the χ^2 test statistic. For each outcome, we approximate (using one million randomizations) exact RI p -values for each of the three two-way comparisons as well as a joint test, which consists of combining the χ^2 test statistics across the three two-way comparisons following Young (2017).⁶ Table 1 displays the results; Tables A2–A5 show the complete data.

As can be seen from Table 1, one out of the 12 p -values for the 12 two-way comparisons is below the pre-registered 5% statistical significance level but this result does not survive adjustment for multiple testing (last column). We also compute an omnibus p -value for all outcomes and treatments in our experiment (by combining all 12 χ^2 test statistics as for the joint tests); this p -value equals 0.578. Study 1 does not provide any evidence that participants’ survey responses are affected by the purported sponsor of our survey.

Table 1: RI p -values Study 1

| | No logo vs. OSU | No logo vs. ND | OSU vs. ND | Joint |
|----------------------|-----------------|----------------|------------|-------|
| Abortion | 0.275 | 0.308 | 0.433 | 0.497 |
| Birth control | 0.546 | 0.988 | 0.650 | 0.826 |
| Religious attendance | 0.903 | 0.963 | 0.968 | 0.995 |
| Prayer | 0.041 | 0.278 | 0.342 | 0.124 |

Note: The table displays RI p -values for the four outcomes (rows) and three two-way comparisons (columns 1–3). The last column shows joint p -values that account for multiple testing within outcomes.

⁵An exact randomization inference balance test (approximated using one million randomizations) of the null hypothesis that the mean vote share of Donald Trump in the 2016 Presidential elections is the same across all three treatment groups yields a p -value of 0.185. For each subject, we looked up county-level vote shares based on the respondent’s IP address. In the interest of full disclosure we note that balance tests were *not* pre-registered.

⁶Specifically, we stack the squared test statistics for each comparison and compute the test statistic $\beta^T \Omega \beta$, where β is the vector of stacked test statistics and Ω is the inverse of the covariance matrix of the stacked test statistics based on 1 million randomly chosen treatment assignments.

STUDY 2

Given the null findings in the first study, we next examined sponsorship effects with questions particularly susceptible to response bias. Response biases are common with unreliable items because they provide multiple interpretations, encouraging respondents to choose the interpretation that places them in the most favorable light. We therefore examined responses to 101-point feeling thermometers which typically exhibit low reliability because respondents cannot make fine-grained distinctions between ratings of, say, 30, 31, or 32 (Krosnick and Presser 2010). Despite their lack of reliability, feeling thermometers are used widely to measure affect toward groups and individuals. We ask respondents about how favorable they are to six groups with clear ideological divisions (liberals, conservatives, Muslims, people on welfare, atheists, and feminists) and paired these questions with banners displaying universities with strong ideological reputations. Respondents were randomly assigned to banners showing no logo, the *Liberty University* logo, and the *UC Berkeley* logo.⁷ Interviewer effects usually bias responses toward the interviewer’s own views (West and Blom 2017). We therefore expect that on the conservative thermometer, individuals in the *Liberty University* condition will tend to offer more favorable ratings than those in the *no logo* condition. And respondents in the *no logo* condition should tend to be more favorable than those in the *UC Berkeley* condition. For the other thermometers, we expect the opposite ordering. For consistency, we therefore reverse-code the conservative thermometer so that all tests expect the highest scores for *UC Berkeley* and the lowest scores for *Liberty University*. Thus, our hypotheses here are one-sided, in contrast to Study 1 and Study 3.

Our target sample size was about 3,000 respondents and our realized sample size $n = 2,339$ (see Table A1). Figure A4 shows the distributions for the six thermometer items by treatment group. Table 2 shows RI p -values. To make our inferences robust to outliers we used the difference in average ranks as our test statistic. We have three two-way comparisons (columns 1–3) for six outcomes (rows), yielding 18 comparisons in total. All tests are *one-sided*; cell entries showing * denote contrasts that failed to have the hypothesized sign. None of the p -values is below the 5% cutoff for statistical significance. The last column shows joint p -values that account for multiple testing within outcomes. None is close to statistical significance. Finally, the omnibus p -value is 0.699; we fail to reject the sharp null hypothesis that any treatment has any effect on any outcome for any participant.

⁷A balance test analogous to the one for study 1 yields a p -value of 0.705.

We also included a three-item self monitoring battery in our survey (measured before exposure to the treatments) because high-self monitors are particularly prone to social desirability bias (Berinsky and Lavine 2012). As stated in our study’s pre-registration, we investigated the one-sided hypothesis that treatment effects would be larger for respondents who score high on the scale than for respondents who score low. The distribution of the self-monitoring scale in our sample is shown in Figure A5; see the figure caption for details on the construction of the scale. Empirically, we compare treatment effects for respondents in the lower tertile to treatment effects for respondents in the upper tertile. Table A6 shows the resulting RI p -values. We find no evidence of treatment effect heterogeneity.

Table 2: RI p -values Study 2

| | No logo vs. UC Berkeley | No logo vs. Liberty U | UC Berkeley vs. Liberty U | Joint |
|-------------------|-------------------------|-----------------------|---------------------------|-------|
| Liberals | * | 0.294 | 0.395 | 0.578 |
| Conservatives | * | 0.208 | 0.276 | 0.440 |
| Muslims | 0.117 | * | 0.427 | 0.271 |
| People on welfare | 0.228 | * | * | 0.473 |
| Atheists | 0.231 | * | * | 0.478 |
| Feminists | 0.331 | * | * | 0.631 |

Note: The table displays RI p -values from one-sided tests for six outcomes (rows) and three two-way comparisons (columns 1–3). The last column shows joint p -values that account for multiple testing within outcomes. * denotes tests in which the test statistic failed to have the hypothesized sign.

STUDY 3

Past work has found interviewer effects for factual knowledge and respondent effort (West and Blom 2017). Some interviewers are better able to encourage respondents to exert the necessary effort to evaluate the survey question, summon their relevant beliefs, and choose the response option that best reflects these beliefs. In online surveys, more prestigious universities may encourage greater effort, but they may also encourage respondents to admit that they do not know the true answer to factual questions. Therefore, in Study 3 we investigated the effect of survey sponsorship on (i) respondents’ willingness to admit not knowing the answers to relatively difficult political knowledge questions and (ii) the amount of effort respondents exerted when answering such questions. Moreover, we also investigated the relationship between survey sponsorship and measures of effort when respondents were asked to answer an open-ended essay question. We randomly assigned sponsor prestige with banners showing no logo, the *Harvard University* logo, and the *Fitchburg State University* logo.⁸

Our target sample size was about 3,000 respondents and our realized sample size $n = 2,212$

⁸A balance test analogous to the ones for the other two studies yields a p -value of 0.859.

(see Table A1). We measured political knowledge using six political knowledge items about (1) U.S. senator’s term length; (2) the budget category on which the U.S. federal government currently spends the least; (3) the current Chief Justice of the U.S. Supreme Court; (4) the job currently held by Theresa May; (5) the current U.S. Attorney General; and (6) the current Secretary of State.⁹ Across these six items, we measured the fraction of “Don’t know” responses as well as the fraction of correct responses. We allow “Don’t know” responses because their use provides a common source of interviewer effects (Krosnick and Presser 2010) and because they are substantively interesting to students of political knowledge (Jessee 2017).

We are interested in whether survey sponsorship affects the fraction of “Don’t know” responses as well as respondents’ effort as measured by the fraction of correct responses and response times. Tables A7–A13 show the distribution of answers to the political knowledge items by treatment group. Tables A14 and A15 show the distribution of the number of “Don’t know” and correct responses by treatment group. Finally, Figure A6 shows the distribution of total response times by treatment group.

We also asked participants to write an open-ended essay.¹⁰ We coded three outcomes: the Flesch reading ease score, the number of characters, and the response time in seconds, all of which we interpret as measures of effort. Figures A7–A9 display the outcome distributions by treatment group.

Table 3 displays RI p -values. Test statistics are absolute differences in ranks (to make inferences robust to outliers) for all outcomes except % *Don’t know* and % *Correct response*, for which we use absolute differences in means. None of the p -values is below 0.05. The omnibus p -value equals 0.839,

⁹The Chief Justice, Theresa May, federal spending, and senate term length questions are variants of items from the 2012 ANES political knowledge battery, chosen because they were some of the most difficult items and because they could be readily extended to four response options plus “Don’t know.” The questions about the Secretary of State and Attorney General were taken from Ahler and Goggin (2017). On March 13, 2018, Donald Trump fired Secretary of State Rex Tillerson, making the question about the current Secretary of State meaningless. After the news broke, we deleted all 94 responses received after March 12, 2018 and replaced this item with a new question which asks which individual is currently a senator from Florida. Exact question wordings and response options are displayed in Tables A7–A13.

¹⁰“In a paragraph, please tell us what being a good citizen means to you.”

providing no evidence against the sharp null hypothesis of no treatment effect for any outcome for any subject.¹¹

Table 3: RI p -values Study 3

| | No logo vs. Harvard | No logo vs. FSU | FSU vs. Harvard | Joint |
|----------------------------------|---------------------|-----------------|-----------------|-------|
| Political knowledge items | | | | |
| % Don't know | 0.404 | 0.802 | 0.560 | 0.694 |
| % Correct response | 0.692 | 0.942 | 0.639 | 0.880 |
| Response time | 0.833 | 0.269 | 0.188 | 0.367 |
| Essay | | | | |
| Number of characters | 0.752 | 0.473 | 0.302 | 0.570 |
| Flesch score | 0.920 | 0.821 | 0.743 | 0.945 |
| Response time | 0.980 | 0.422 | 0.436 | 0.659 |

Note: The table displays RI p -values for six outcomes (rows) and three two-way comparisons (columns 1–3). The last column shows joint p -values that account for multiple testing within outcomes.

CONCLUSION AND IMPLICATIONS

Surveys conducted on MTurk routinely include university logos in banner images. Contrary to a large literature demonstrating that even seemingly minute alterations of the survey experience can induce changes in survey responses, we found no evidence that their prominent display affects respondents' reported religiosity and policy attitudes (Study 1), group affect (Study 2), or political knowledge and effort levels (Study 3). Since we cannot feasibly study a representative sample of universities or survey items, it is of course logically possible that sponsorship effects could exist for other universities or other survey items. Nonetheless, we examined the effect of logos for some of the most recognizable universities using items previously linked to various forms of response bias. If logo-induced sponsorship effects were widespread, we would expect some systematic differences between our treatment groups. Likewise, one might imagine larger effects for regionally-proximate respondents, alumni, or other subsets of individuals who would be particularly sensitive to such cues. Although possible, our results show that these groups are not sufficiently large portions of typical MTurk samples to lead to detectable sponsorship effects in our experiments (and, presumably, in other social science research using MTurk samples). Further, we show in Study 2 that even high self-monitors — a group prone to socially-desirable responses — fail to exhibit sensitivity to

¹¹This null finding cannot simply be explained by participants not paying attention to the banners. Table A16 shows that 58% of participants in the Harvard group recall seeing the Harvard banner at the end of the survey. The corresponding number for respondents in the Fitchburg State University group is 48%.

university logos. For these reasons, we believe that including such logos constitutes little threat to the validity of survey responses, validating their common use in social science research using online samples.

REFERENCES

- Ahler, Douglas J. and Stephen N. Goggin. 2017. "Assessing Political Knowledge: Problems and Solutions in Online Surveys." Working Paper.
- Berinsky, Adam J. and Howard Lavine. 2012. Self-Monitoring and Political Attitudes. In *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*, ed. John Aldrich and Kathleen M. McGraw. Princeton, N.J.: Princeton University Press pp. 27–45.
- Burdein, Inna, Milton Lodge and Charles Taber. 2006. "Experiments on the Automaticity of Political Beliefs and Attitudes." *Political Psychology* 27(3):359–371.
- Clifford, Scott, Ryan M. Jewell and Philip D. Waggoner. 2015. "Are samples drawn from Mechanical Turk valid for research on political ideology?" *Research & Politics* 2(4):1–9.
- Connors, Elizabeth C., Yanna Krupnikov and John Barry Ryan. 2017. "How Transparency Affects Survey Responses." Working paper.
- Cotter, Patrick R., Jeffrey Cohen and Philip B. Coulter. 1982. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly* 46(2):278–284.
- Druckman, James N. and Thomas J. Leeper. 2012. "Learning More from Political Communication Experiments: Pretreatment and Its Effects." *American Journal of Political Science* 56(4):875–896.
- Edwards, Michelle L., Don A. Dillman and Jolene D. Smyth. 2014. "An Experimental Test of the Effects of Survey Sponsorship on Internet and Mail Survey Response." *Public Opinion Quarterly* 78(3):734–750.
- Fisher, Ronald A. 1935. *The Design of Experiments*. Edinburgh; London: Oliver and Boyd.
- Galesic, Mirta, Roger Tourangeau, Mick P. Couper and Frederick G. Conrad. 2008. "Eye-Tracking Data: New Insights on Response Order Effects and Other Cognitive Shortcuts in Survey Responding." *Public Opinion Quarterly* 72(5):892–913.
- Imbens, Guido W. and Donald B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

- Jessee, Stephen A. 2017. ““Don’t Know” Responses, Personality, and the Measurement of Political Knowledge.” *Political Science Research and Methods* 5(4):711–731.
- Kinder, Donald R. and Lynn M. Sanders. 1990. “Mimicking Political Debate with Survey Questions: The Case of White Opinion on Affirmative Action for Blacks.” *Social Cognition* 8(1):73–103.
- Krosnick, Jon A. and Stanley Presser. 2010. Question and Questionnaire Design. In *Handbook of Survey Research*, ed. Peter V. Marsden and James D. Wright. Bingley, UK: Emerald Group Publishing pp. 263–314.
- Krupnikov, Yanna and Adam Seth Levine. 2014. “Cross-Sample Comparisons and External Validity.” *Journal of Experimental Political Science* 1(1):59–80.
- Leeper, Thomas J. and Emily Thorson. 2015. Minimal Sponsorship-Induced Bias in Web Survey Data. Working paper.
- Mummolo, Jonathan and Erik Peterson. 2017. “Demand Effects in Survey Experiments: An Empirical Assessment.” Working paper.
- Sigelman, Lee. 1981. “Question-Order Effects on Presidential Popularity.” *Public Opinion Quarterly* 45(2):199–207.
- Tourangeau, Roger, Robert M. 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(3):299–321.
- West, Brady T. and Annelies G. Blom. 2017. “Explaining Interviewer Effects: A Research Synthesis.” *Journal of Survey Statistics and Methodology* 5(2):175–211.
- White, Ariel, Anton Strezhnev, Christopher Lucas, Dominika Kruszewska and Connor Huff. 2018. “Investigator Characteristics and Respondent Behavior in Online Surveys.” *Journal of Experimental Political Science* 5(1):56–67.
- Young, Alwyn. 2017. “Channelling Fisher: Randomization tests and the statistical insignificance of seemingly significant experimental result.” Working paper.

Supplemental Information

[to be published online]

LIST OF FIGURES

| | | |
|----|---|--------|
| A1 | Screenshots from Study 1 | SI-P1 |
| A2 | Screenshots from Study 2 | SI-P2 |
| A3 | Screenshots from Study 3 | SI-P3 |
| A4 | Study 2 Feeling thermometers | SI-P20 |
| A5 | Study 2: Self-monitoring scale | SI-P21 |
| A6 | Study 3: Response times for political knowledge items | SI-P22 |
| A7 | Study 3: Response times for essay question | SI-P23 |
| A8 | Study 3: Flesch scores for essay question | SI-P24 |
| A9 | Study 3: Number of characters in essay responses | SI-P25 |

LIST OF TABLES

| | | |
|-----|--|--------|
| A1 | Target and realized sample sizes for studies 1–3 | SI-P4 |
| A2 | Study 1 — Abortion | SI-P5 |
| A3 | Study 1 — Birth control | SI-P6 |
| A4 | Study 1 — Religious attendance | SI-P7 |
| A5 | Study 1 — Prayer | SI-P8 |
| A6 | RI <i>p</i> -values Study 2: Treatment effect heterogeneity | SI-P9 |
| A7 | Study 3: Political knowledge question: “What is the term, in years, of a U.S. senator?” | SI-P10 |
| A8 | Study 3: Political knowledge question: “Who is currently the U.S. Secretary of State?” | SI-P11 |
| A9 | Study 3: Political knowledge question: “Which of these people is a U.S. Senator from Florida?” | SI-P12 |
| A10 | Study 3: Political knowledge question: “On which of the following does the U.S. federal government currently spend the least?” | SI-P13 |
| A11 | Study 3: Political knowledge question: “Who is the current Chief Justice of the United States?” | SI-P14 |
| A12 | Study 3: Political knowledge question: “What job or political office does Theresa May currently hold?” | SI-P15 |
| A13 | Study 3: “Who is currently the U.S. Attorney General?” | SI-P16 |
| A14 | Study 3: Distribution of outcome variable by treatment group: Number of Don’t knows | SI-P17 |
| A15 | Study 3: Distribution of outcome variable by treatment group: Correct answers . . | SI-P18 |
| A16 | Study 3: Recall of banner by treatment group | SI-P19 |

How often do you attend religious services?

▼

- Never
- Less than once a year
- About once or twice a year
- Several times a year
- About once a month
- 2-3 times a month
- Nearly every week
- Every week
- Several times a week
- Don't know



THE OHIO STATE UNIVERSITY

How often do you attend religious services?

▼

- Never
- Less than once a year
- About once or twice a year
- Several times a year
- About once a month
- 2-3 times a month
- Nearly every week
- Every week
- Several times a week
- Don't know



Powered by Qualtrics



UNIVERSITY OF NOTRE DAME

How often do you attend religious services?

▼

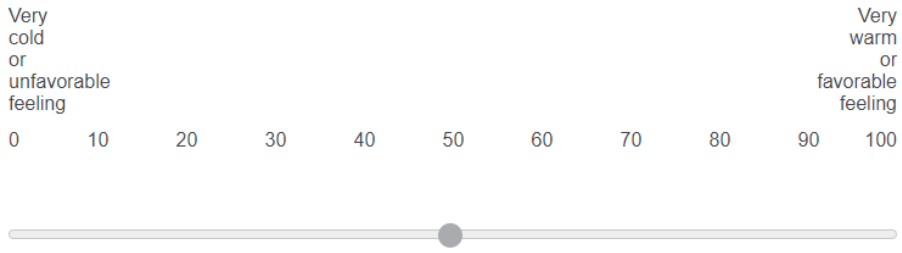
- Never
- Less than once a year
- About once or twice a year
- Several times a year
- About once a month
- 2-3 times a month
- Nearly every week
- Every week
- Several times a week
- Don't know



Powered by Qualtrics

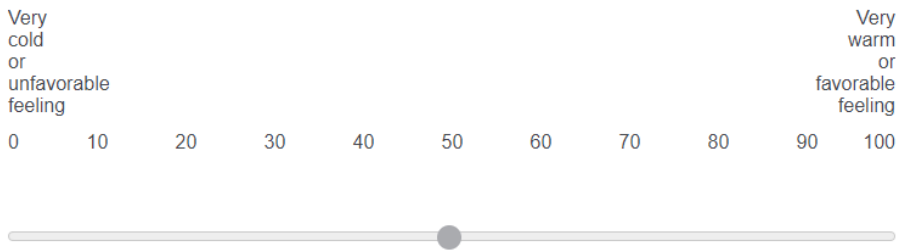
Figure A1: Screenshots from Study 1

Conservatives



Berkeley
UNIVERSITY OF CALIFORNIA

Conservatives



LIBERTY
UNIVERSITY

Conservatives

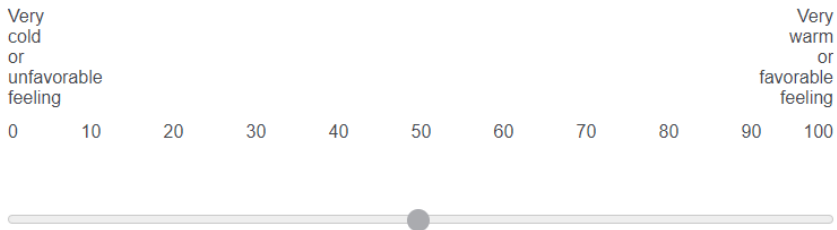


Figure A2: Screenshots from Study 2

Who is currently the U.S. Secretary of State?

- John Kerry
- Kellyanne Conway
- Rex W. Tillerson
- Mike Pompeo
- Don't know



Who is currently the U.S. Secretary of State?

- Mike Pompeo
- John Kerry
- Kellyanne Conway
- Rex W. Tillerson
- Don't know



Who is currently the U.S. Secretary of State?

- Rex W. Tillerson
- John Kerry
- Mike Pompeo
- Kellyanne Conway
- Don't know

Figure A3: Screenshots from Study 3

Table A1: Target and realized sample sizes for studies 1–3

| | Study 1 | Study 2 | Study 3 |
|---|----------------|----------------|----------------|
| Target sample size | ≈ 2000 | ≈ 3000 | ≈ 3000 |
| Respondents who were located outside 50 U.S. states | 46 | 84 | 125 |
| Respondents who failed to consent to survey participation | 32 | 47 | 76 |
| Respondents who failed to complete the survey | 7 | 22 | 218 |
| Respondents with lat/long coordinates identical to previous responses | 322 | 592 | 585 |
| Realized sample size | 1646 | 2339 | 2212 |

Table A2: Study 1 — Abortion

| | No logo | OSU | ND |
|---|---------|-----|-----|
| By law, a woman should always be able to obtain an abortion as a matter of personal choice. | 302 | 298 | 322 |
| The law should permit abortion for reasons other than rape, incest, or danger to the woman. | 62 | 74 | 72 |
| The law should permit abortion only in case of rape, incest, or when the woman's life is in danger. | 137 | 109 | 126 |
| By law, abortion should never be permitted. | 42 | 44 | 31 |
| Don't know. | 6 | 10 | 11 |

Note: The table displays the number of respondents in each response category for each treatment. OSU = Ohio State University. ND = University of Notre Dame. Question wording: "There has been some discussion of abortion in recent years. Which one of the opinions on this page best agrees with your view?"

Table A3: Study 1 — Birth control

| | No logo | OSU | ND |
|-------------------|---------|-----|-----|
| Strongly agree | 222 | 189 | 223 |
| Agree | 186 | 195 | 187 |
| Disagree | 65 | 68 | 72 |
| Strongly disagree | 54 | 58 | 56 |
| Don't know | 22 | 25 | 24 |

Note: The table displays the number of respondents in each response category for each treatment. OSU = Ohio State University. ND = University of Notre Dame. Question wording: ‘Do you strongly agree, agree, disagree, or strongly disagree that methods of birth control should be available to teenagers between the ages of 14 and 16 if their parents do not approve?’

Table A4: Study 1 — Religious attendance

| | No logo | OSU | ND |
|----------------------------|---------|-----|-----|
| Several times a week | 11 | 9 | 9 |
| Every week | 43 | 37 | 41 |
| Nearly every week | 21 | 16 | 22 |
| 2-3 times a month | 21 | 24 | 30 |
| About once a month | 27 | 17 | 21 |
| Several times a year | 49 | 46 | 53 |
| About once or twice a year | 62 | 57 | 64 |
| Less than once a year | 84 | 88 | 89 |
| Never | 230 | 240 | 231 |
| Don't know | 1 | 1 | 2 |

Note: The table displays the number of respondents in each response category for each treatment. OSU = Ohio State University. ND = University of Notre Dame. Question wording: “How often do you attend religious services?”

Table A5: Study 1 — Prayer

| | No logo | OSU | ND |
|-----------------------|---------|-----|-----|
| Never | 188 | 214 | 208 |
| Less than once a week | 93 | 91 | 94 |
| Once a week | 27 | 18 | 36 |
| Several times a week | 53 | 71 | 69 |
| Once a day | 83 | 58 | 68 |
| Several times a day | 93 | 75 | 76 |
| Don't know | 12 | 8 | 11 |

Note: The table displays the number of respondents in each response category for each treatment. OSU = Ohio State University. ND = University of Notre Dame. Question wording: “About how often do you pray?”

Table A6: RI p -values Study 2: Treatment effect heterogeneity

| | No logo vs. UC Berkeley | No logo vs. Liberty U | UC Berkeley vs. Liberty U | Joint |
|-------------------|-------------------------|-----------------------|---------------------------|-------|
| Liberals | * | 0.364 | * | 0.681 |
| Conservatives | * | * | * | * |
| Muslims | * | 0.123 | * | 0.281 |
| People on welfare | * | 0.159 | * | 0.353 |
| Atheists | 0.361 | * | * | 0.676 |
| Feminists | * | 0.154 | * | 0.343 |

Note: The table displays RI p -values from one-sided tests for treatment effect heterogeneity for six outcomes (rows) and three two-way comparisons (columns 1–3). The last column shows joint p -values that account for multiple testing within outcomes. * denotes tests in which the test statistic failed to have the hypothesized sign.

Table A7: Study 3: Political knowledge question: “What is the term, in years, of a U.S. senator?”

| | No logo | Harvard | Fitchburg State |
|------------|---------|---------|-----------------|
| 2 years | 124 | 147 | 145 |
| 4 years | 124 | 144 | 133 |
| 6 years | 357 | 359 | 359 |
| 8 years | 59 | 54 | 65 |
| Don't know | 43 | 50 | 49 |

Note: The table displays the number of respondents in each response category for each treatment.

Table A8: Study 3: Political knowledge question: “Who is currently the U.S. Secretary of State?”

| | No logo | Harvard | Fitchburg State |
|------------------|---------|---------|-----------------|
| John Kerry | 48 | 37 | 43 |
| Kellyanne Conway | 30 | 23 | 27 |
| Mike Pompeo | 25 | 34 | 17 |
| Rex W. Tillerson | 244 | 254 | 258 |
| Don't know | 65 | 92 | 75 |

Note: The table displays the number of respondents in each response category for each treatment. Note that only a subset of the sample was asked this question.

Table A9: Study 3: Political knowledge question: “Which of these people is a U.S. Senator from Florida?”

| | No logo | Harvard | Fitchburg State |
|-------------|---------|---------|-----------------|
| Jeb Bush | 37 | 42 | 45 |
| Marco Rubio | 179 | 193 | 194 |
| Rand Paul | 6 | 8 | 11 |
| Ted Cruz | 35 | 33 | 34 |
| Don't know | 38 | 38 | 47 |

Note: The table displays the number of respondents in each response category for each treatment. Note that only a subset of the sample was asked this question.

Table A10: Study 3: Political knowledge question: “On which of the following does the U.S. federal government currently spend the least?”

| | No logo | Harvard | Fitchburg State |
|------------------|---------|---------|-----------------|
| Foreign aid | 234 | 226 | 243 |
| Medicare | 135 | 156 | 152 |
| National defense | 48 | 46 | 36 |
| Social security | 161 | 182 | 187 |
| Don't know | 129 | 144 | 133 |

Note: The table displays the number of respondents in each response category for each treatment.

Table A11: Study 3: Political knowledge question: “Who is the current Chief Justice of the United States?”

| | No logo | Harvard | Fitchburg State |
|---------------------|---------|---------|-----------------|
| John Roberts | 373 | 399 | 406 |
| Paul Ryan | 52 | 29 | 44 |
| Robert Mueller | 68 | 82 | 79 |
| Ruth Bader Ginsburg | 77 | 88 | 68 |
| Don't know | 137 | 156 | 154 |

Note: The table displays the number of respondents in each response category for each treatment.

Table A12: Study 3: Political knowledge question: “What job or political office does Theresa May currently hold?”

| | No logo | Harvard | Fitchburg State |
|--|---------|---------|-----------------|
| Chancellor of Germany | 46 | 50 | 35 |
| Prime Minister of the United Kingdom | 401 | 422 | 436 |
| Secretary-General of the United Nations | 35 | 34 | 38 |
| Speaker of the U.S. House of Representatives | 32 | 29 | 41 |
| Don't know | 193 | 219 | 201 |

Note: The table displays the number of respondents in each response category for each treatment.

Table A13: Study 3: “Who is currently the U.S. Attorney General?”

| | No logo | Harvard | Fitchburg State |
|---------------|---------|---------|-----------------|
| Jeff Sessions | 504 | 562 | 544 |
| Loretta Lynch | 22 | 20 | 25 |
| Mike Pence | 32 | 28 | 36 |
| Rick Perry | 33 | 27 | 25 |
| Don't know | 116 | 117 | 121 |

Note: The table displays the number of respondents in each response category for each treatment.

Table A14: Study 3: Distribution of outcome variable by treatment group: Number of Don't knows

| | No logo | Harvard | Fitchburg State |
|---|---------|---------|-----------------|
| 0 | 224 | 227 | 238 |
| 1 | 83 | 91 | 79 |
| 2 | 47 | 47 | 44 |
| 3 | 19 | 33 | 21 |
| 4 | 21 | 16 | 21 |
| 5 | 8 | 19 | 14 |
| 6 | 10 | 7 | 3 |

Note: By treatment group, the table displays the number of respondents with a given number of “Don't know” responses to the six political knowledge items.

Table A15: Study 3: Distribution of outcome variable by treatment group: Correct answers

| | No logo | Harvard | Fitchburg State |
|---|---------|---------|-----------------|
| 0 | 63 | 70 | 80 |
| 1 | 108 | 92 | 94 |
| 2 | 96 | 115 | 88 |
| 3 | 102 | 125 | 124 |
| 4 | 105 | 132 | 128 |
| 5 | 132 | 130 | 136 |
| 6 | 101 | 90 | 101 |

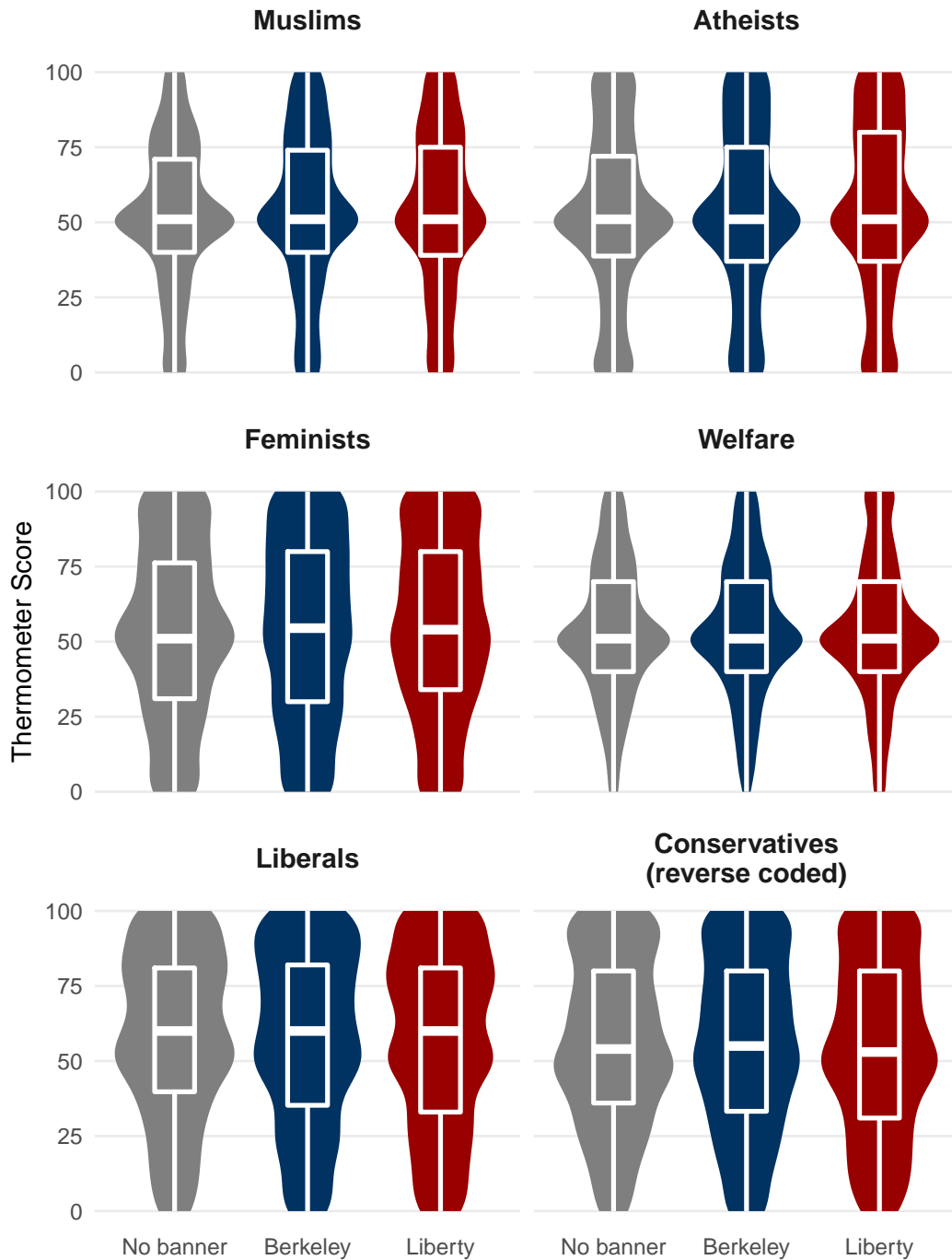
Note: By treatment group, the table displays the number of respondents with a given number of correct responses to the six political knowledge items. Don't knows are coded as incorrect.

Table A16: Study 3: Recall of banner by treatment group

| | No logo | Harvard | Fitchburg State |
|---|---------|---------|-----------------|
| I do not remember a logo | 626 | 168 | 144 |
| Harvard University | 2 | 437 | 6 |
| Fitchburg State University | 13 | 6 | 363 |
| I remember a logo, but I do not remember the organization | 47 | 123 | 203 |
| Brookings Institution | 4 | 2 | 2 |
| Cato Institute | 0 | 2 | 2 |
| Gallup | 3 | 5 | 1 |
| Massachusetts Institute of Technology (MIT) | 5 | 7 | 15 |
| Pew Research Center | 5 | 1 | 5 |
| Salem State University | 2 | 3 | 10 |
| % correct recall | 89 | 58 | 48 |

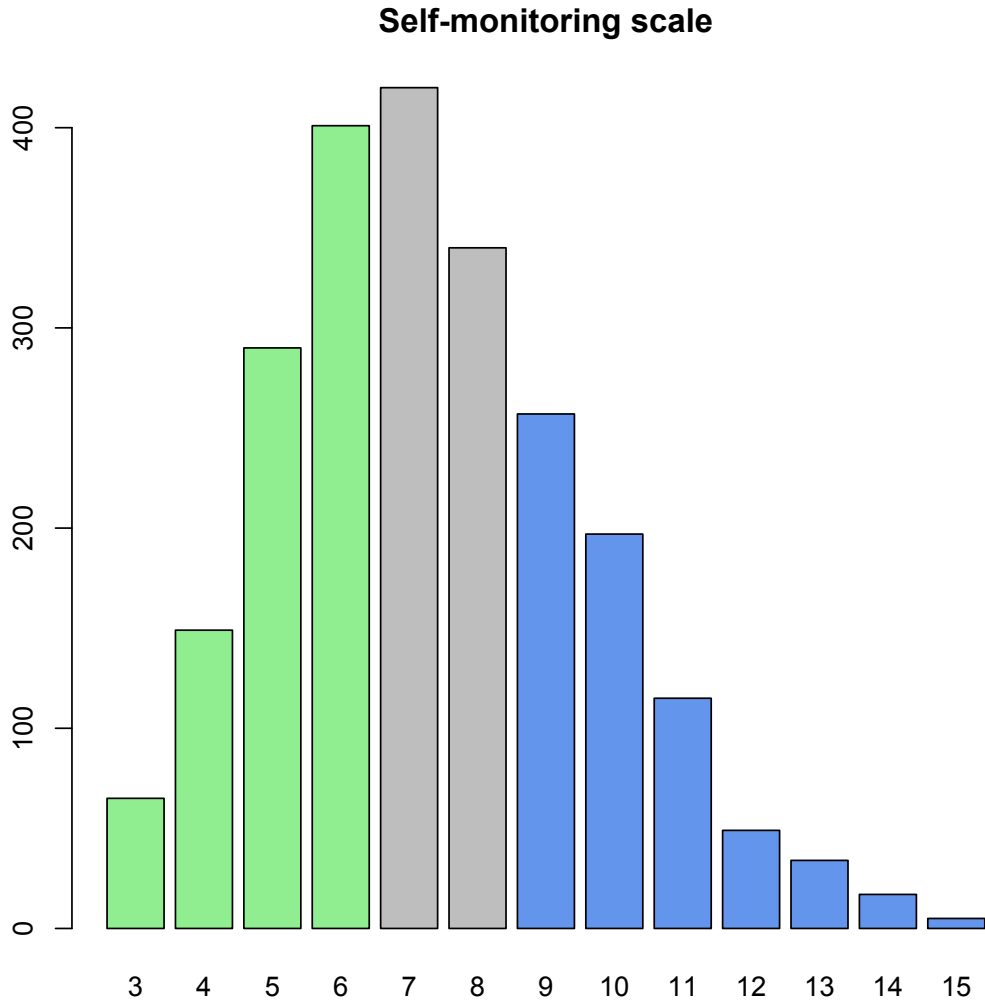
Note: By treatment group, the table displays the number of respondents falling into each response category when prompted with “A logo may have been displayed at the top of the last few pages. If so, do you recall what organization the logo represents?” after the end of the survey.

Figure A4: Study 2 Feeling thermometers



Note: The figure summarizes the feeling thermometer scores with box plots indicating the 25th, 50th, and 75th percentiles on top of probability density distributions. Greater values indicate warmer feelings toward the group, except the conservative thermometer, which has been reverse coded so that greater values indicate cooler feelings toward conservatives.

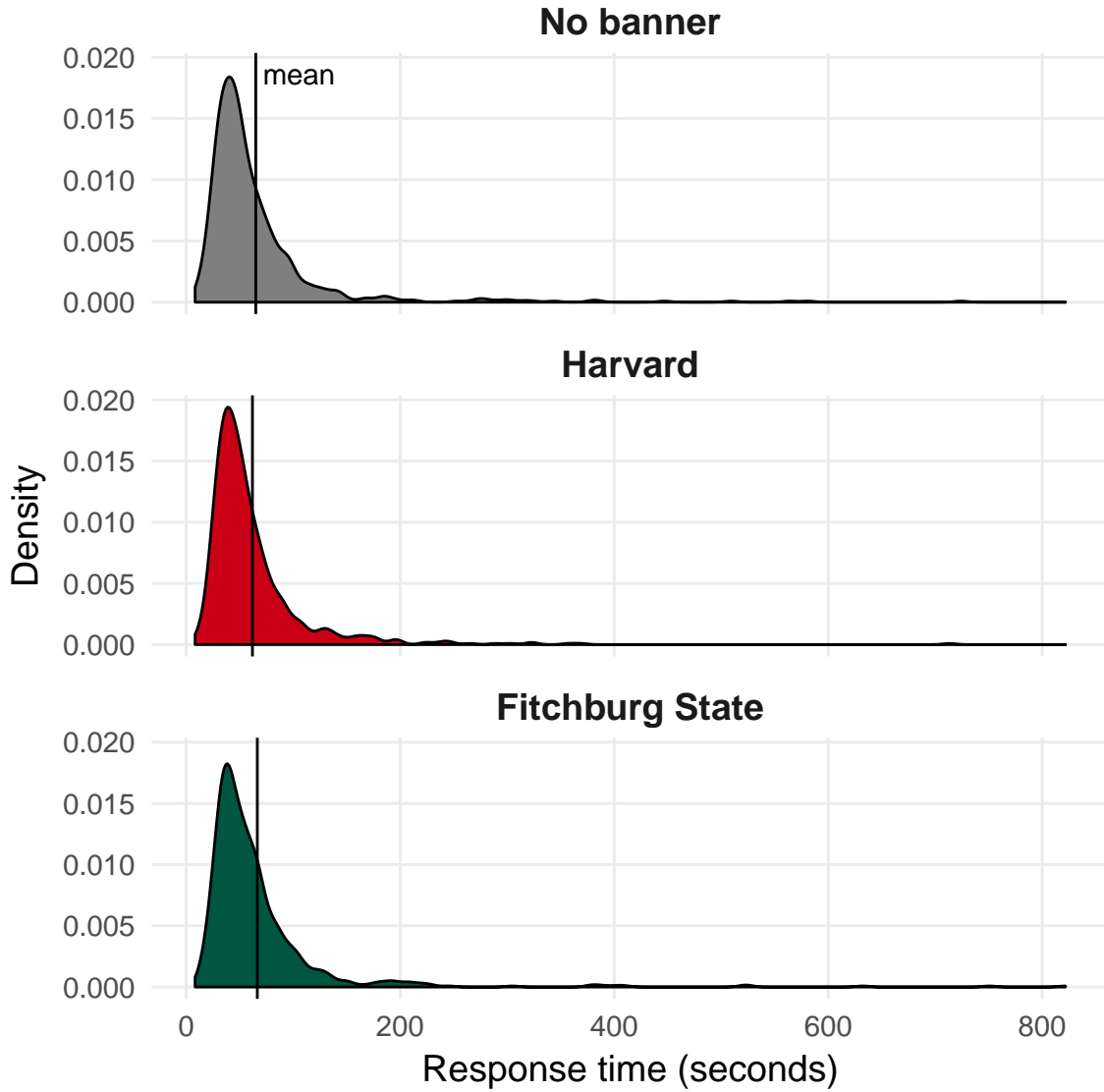
Figure A5: Study 2: Self-monitoring scale



Note: The figure shows the sample distribution of the self-monitoring scale. Colors denote the sample tertiles. The scale is constructed additively from three pre-treatment survey questions:

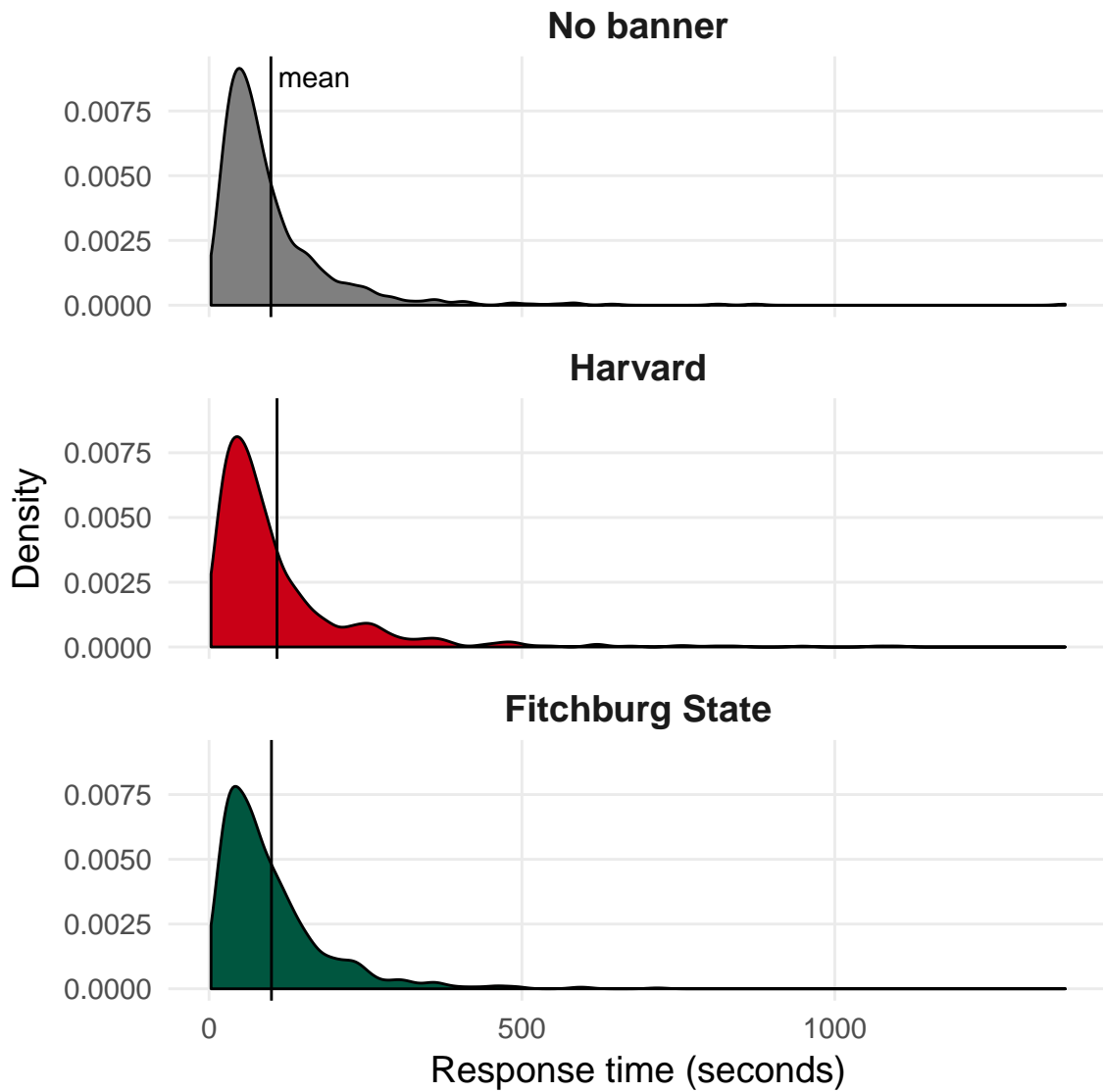
1. “When you’re with other people, how often do you put on a show to impress or entertain them?” Response options are “Always” (5), “Most of the time” (4), “About half the time” (3), “Once in a while” (2), and “Never” (1).
2. “How good or bad of an actor would you be?” Response options are “Excellent” (5), “Good” (4), “Fair” (3), “Poor” (2), and “Very poor” (1).
3. “When you’re in a group of people, how often are you the center of attention?” Response options are “Always” (5), “Most of the time” (4), “About half the time” (3), “Once in a while” (2), and “Never” (1).

Figure A6: Study 3: Response times for political knowledge items



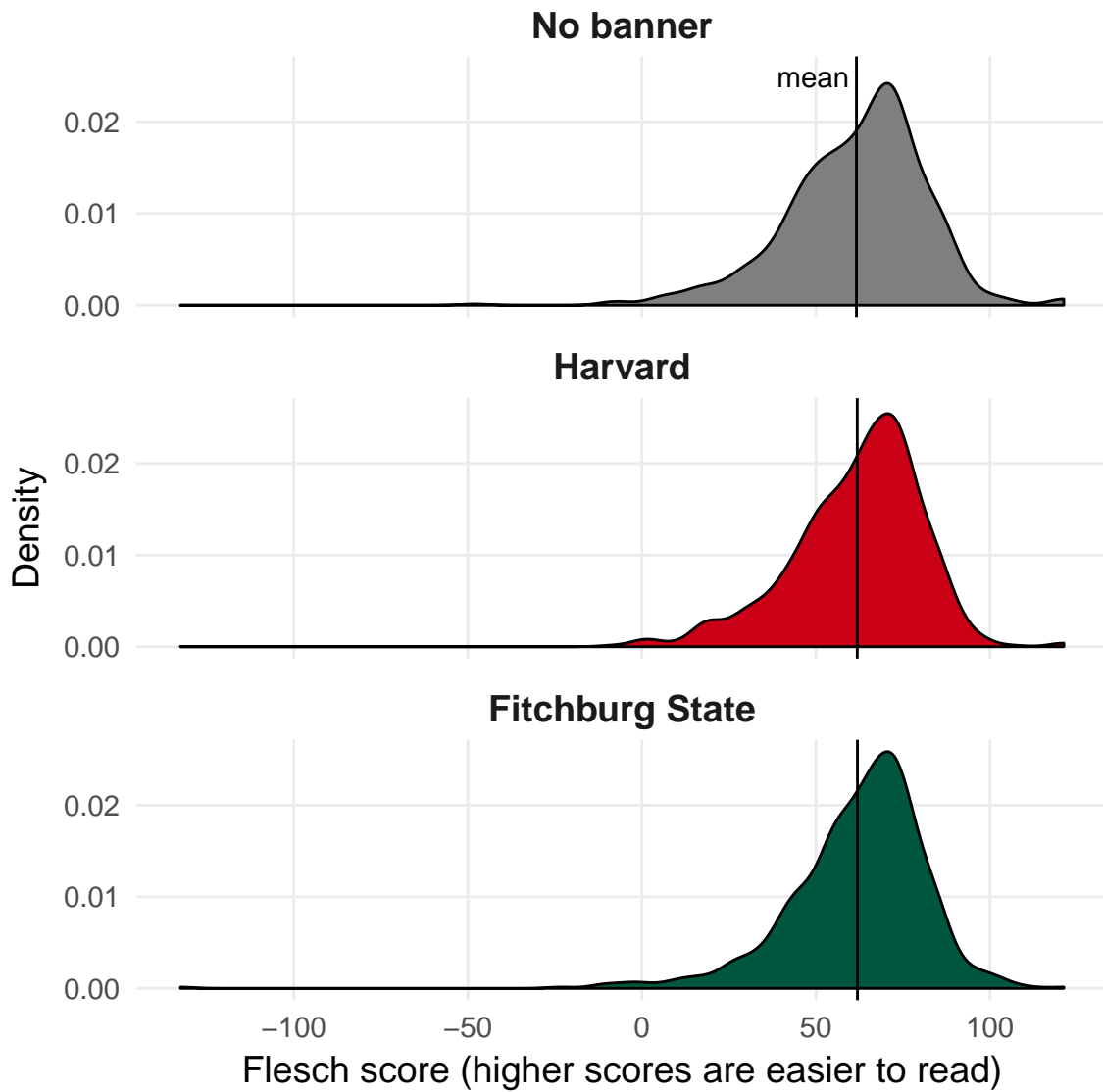
Note: The plots show the sample distribution of response times summed over all political knowledge items by treatment group.

Figure A7: Study 3: Response times for essay question



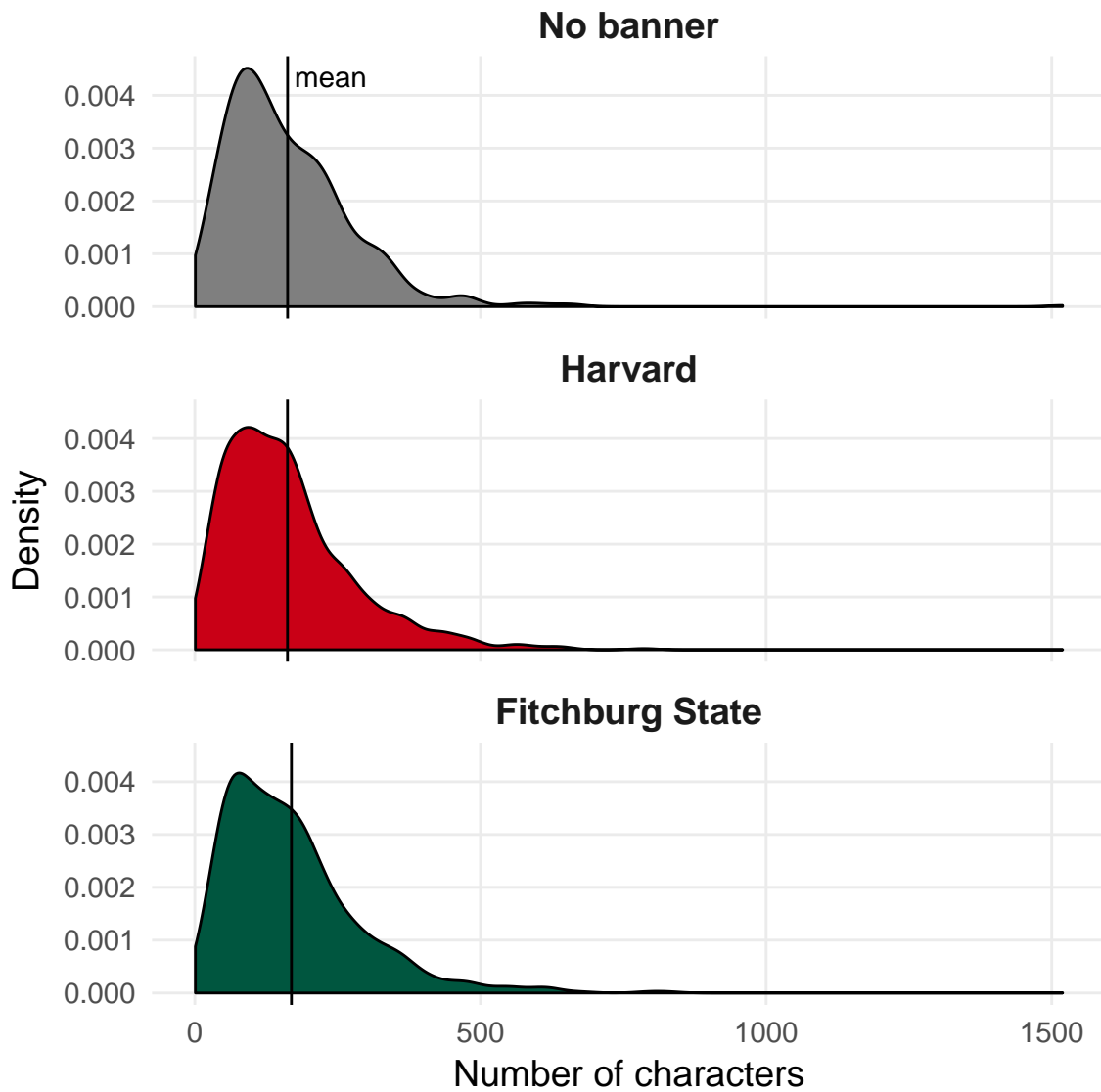
Note: The figure shows the sample distribution of response times for the essay question by treatment group.

Figure A8: Study 3: Flesch scores for essay question



Note: The figure shows the sample distribution of Flesch reading ease scores for the essay question by treatment group.

Figure A9: Study 3: Number of characters in essay responses



Note: The figure shows the sample distribution of the number of characters written by respondents, by treatment group.