

# Investigator Characteristics and Respondent Behavior in Online Surveys

Ariel White\*, Anton Strezhnev†, Christopher Lucas†, Dominika Kruszewska† and Connor Huff†

## Abstract

Prior research demonstrates that responses to surveys can vary depending on the race, gender, or ethnicity of the investigator asking the question. We build upon this research by empirically testing how information about researcher identity in *online surveys* affects subject responses. We do so by conducting an experiment on Amazon's Mechanical Turk in which we vary the name of the researcher in the advertisement for the experiment and on the informed consent page in order to cue different racial and gender identities. We fail to reject the null hypothesis that there is no difference in how respondents answer questions when assigned to a putatively black/white or male/female researcher.

**Keywords:** Gender, race, online surveys, social desirability bias, demand effects, Mechanical Turk

## INTRODUCTION

Researchers conducting in-person and telephone surveys have long found that the ways respondents answer questions can vary depending on the race, gender, or ethnicity of the interviewer (Adida et al., 2016; Cotter et al., 1982; Davis, 1997; Davis and Silver, 2003; Hatchett and Schuman, 1975; Huddy et al., 1997;

Authors are listed in reverse-alphabetical order and contributed equally. For helpful comments, we thank Matthew Blackwell, Joshua Kertzer, Gary King, Jonathan Ladd, Robert Schub, Dustin Tingley, and participants of the Harvard Experimental Political Science Graduate Student Conference and MPSA. For generously funding this research, we thank the Harvard Experiments Working Group, the Harvard Center for American Political Studies, and the Multidisciplinary Program in Inequality and Social Policy at the Harvard Kennedy School. This study was reviewed by Harvard University's IRB with protocol #IRB15-2260 on July 1, 2015. It was pre-registered on EGAP with <http://egap.org/registration/15011D20151014AB>. The data and code required to replicate all analyses in this article can be found on Dataverse (DOI:10.7910/DVN/R8PNCP).

\*Massachusetts Institute of Technology, Cambridge, MA 02142, USA, email: [arwhi@mit.edu](mailto:arwhi@mit.edu)

†Department of Government, Harvard University, Cambridge, MA 02138, USA, email: [astrezhnev@fas.harvard.edu](mailto:astrezhnev@fas.harvard.edu), [clucas@fas.harvard.edu](mailto:clucas@fas.harvard.edu), [dkruszewska@fas.harvard.edu](mailto:dkruszewska@fas.harvard.edu), [cdezzanihuff@fas.harvard.edu](mailto:cdezzanihuff@fas.harvard.edu)

Reese et al., 1986). This is generally argued to occur for two main reasons. First, the provision of information about the investigator could create demand effects whereby the subjects guess the purpose of the study or the interviewer's views and change their responses to align with this perceived purpose.<sup>1</sup> Second, potential subjects may be more or less comfortable answering questions from researchers with a particular identity and subsequently either refuse to participate in studies, decline answering certain questions, or censor the ways in which they answer, all of which could substantively change the results of survey research. Researchers often seek to mitigate these concerns when designing surveys.<sup>2</sup>

In this paper, we build upon prior research and empirically test whether researcher identity affects survey responses in *online survey platforms*. We do so by varying information about the researcher—conveyed through their name—in both the advertisement for survey participation and the informed consent page. We take this approach for two reasons. First, the inclusion of researcher names at each of these junctures is a common practice. Second, an emerging strain of research throughout the social sciences demonstrates how inferences made from names can affect behavioral outcomes even in the absence of in-person or telephone interactions.<sup>3</sup> We go on to test how this variation in the researcher name affects the ways in which respondents answer questions in online surveys.

The experiment conducted in this paper contributes to an expanding strain of research exploring the composition and attributes of online survey pools.<sup>4</sup> Our findings help to interpret the substantive results of prior studies that used online surveys,<sup>5</sup> and also provide guidelines for researchers as they move forward. In this study, we fail to reject the null hypothesis of no difference in respondents' behavior when assigned to a putatively black/white or female/male researcher. Our estimates suggest that there could be a substantively small difference between question responses for putatively male and female researchers, but given the high

<sup>1</sup>We can consider concerns about social desirability bias as falling into this category.

<sup>2</sup>For example, Grewal and Ritchie (2006), Schaeffer et al. (2010), and Survey Research Center (2010) explicitly advise researchers to consider interviewer effects as part of the research design, though more recent research demonstrates that demand effect concerns might be overstated for survey experiments (Mummolo and Peterson, 2017). See also Berrens et al. (2003) for a discussion of the advantage of internet surveys in reducing interviewer bias compared to telephone or in-person surveys, and Bush and Prather (2017) for how the mode of technology used to conduct surveys can substantively affect survey responses.

<sup>3</sup>See, for example, Bertrand and Mullainathan (2004), Butler and Broockman (2011), White et al. (2015), Einstein and Glick (2017), Edelman et al. (2017), and most recently Butler and Homola (2017).

<sup>4</sup>See, for example, Berinsky et al. (2014), Chandler et al. (2014), Krupnikov and Levine (2014), Clifford et al. (2015), Huff and Tingley (2015), Mullinix et al. (2015), Levay et al. (2016), Leeper and Thorson (2015).

<sup>5</sup>A few prominent examples of political science articles using online samples drawn from Mechanical Turk have been published in the *American Political Science Review* (Tomz and Weeks, 2013), *American Journal of Political Science* (Healy and Lenz, 2014), *Comparative Political Studies* (Charnysh et al., 2014), *International Organization* (Wallace, 2013), and *the Journal of Conflict Resolution* (Kriner and Shen, 2013).

*Table 1*  
**Names Used for Each of the Four Investigator Name Manipulations**

Black men	Deshawn Booker	Tyrone Robinson
Black women	Ebony Gaines	Deja Washington
White men	Connor Schroeder	Brett Walsh
White women	Molly Ryan	Laurie Yoder

Investigator name manipulations are based on lists from Bertrand and Mullainathan (2004), Fryer, Jr. and Levitt (2004), and Word et al. (2008).

power of the experiment, we are able to bound the substantive size of the effect. We conclude that these differences are likely substantively negligible for most researchers. In general, the results of this paper demonstrate that researchers need not worry that using their own names in either survey advertisements or online consent forms will substantively affect online survey results.

## EXPERIMENTAL DESIGN

In the experiment, each respondent was treated with one researcher name intended to cue race and gender, appearing first in the advertisement for the survey and then in the consent form inside the survey. The experiment was conducted on Amazon's Mechanical Turk (MTurk), where it is common for researchers' names to appear at both of these points.<sup>6</sup> To generate the names associated with each of these manipulations, we combined three commonly used lists of racially distinct first and last names.<sup>7</sup> We crossed the lists of first and last names to produce many possible combinations<sup>8</sup> and drew two names for each of the four manipulation categories (black men, black women, white men, and white women). The full list of names used in this experiment is presented in Table 1.

We then created accounts under the names of our hypothetical researchers ("Ebony Gaines," "Brett Walsh," etc.) and recruited subjects through these named

<sup>6</sup>Readers will note that this design captures two stages: first, selection into the survey, and second, the ways respondents answer questions conditional on having selected into the survey. In Section J of the Online Appendix, we present results from a different experiment in which we randomize the name of the researcher *only on the consent form* with a generic account name. Doing so allows us to estimate the effect of varying the researcher name only in the consent form where there is no initial selection step. The results from this second experiment are substantively consistent with what we present in the remainder of this paper.

<sup>7</sup>First names were drawn from a combination of lists found in Bertrand and Mullainathan (2004) and Fryer, Jr. and Levitt (2004), while last names were drawn from lists in Word et al. (2008) and Bertrand and Mullainathan (2004). Our instrument did not include a manipulation check but the studies from which we drew the list of names have shown that they are racially-distinctive enough for respondents to make inferences about the person's racial identity. We are thus confident that the names we used were highly informative about the race and gender of the individual conducting the study.

<sup>8</sup>We omitted a few randomly-generated names that already belonged to celebrities, such as Jermaine Jackson.

*Table 2*  
**The Number of Unique Accounts on MTurk Using Real Names**

Search term	Accounts using real names	Total accounts	Proportion using name
Politics	3	5	0.6
Survey	121	169	0.72
Research	21	41	0.51
Academic	5	9	0.56
University	1	4	0.25
Political science	1	1	1
Psychology	23	31	0.74
Economics	1	1	1
Total	176	261	0.67

To calculate these amounts, we searched for the specified term and then scraped all account names on August 15, 2016. Next, we manually classified all unique account names as either a real identifiable name or any other naming scheme (lab name, nonsensical string, etc.).

accounts. We also included these researcher names on the consent forms for our study. This dual approach is both realistic and methodologically useful. Many Institutional Review Boards require that the researcher include their names on the consent form, and as shown in Table 2, a large number of researchers post studies on platforms such as MTurk under their own names. Given these practices, the substantive nature of treatment is consistent with common practices for researchers using the MTurk survey pool. Moreover, the research design allows us to measure how knowledge about researchers' identities can shape not only the nature of responses, but the overall response rate.<sup>9</sup> Posting the survey from named researcher accounts means that potential respondents see the name of the researcher before deciding whether or not to participate, allowing us to capture the selection process that may occur in real studies.

However, including the treatment in the recruitment process poses design challenges. We could not simply post all treatment conditions simultaneously, because users would then see eight identical surveys posted under eight different researcher names and immediately understand the purpose of the experiment. Instead, we set up the experiment such that any user could only observe one treatment condition by pre-recruiting a pool of respondents.

First, we ran a pre-survey asking only one question<sup>10</sup> that captured the unique MTurk "workerID" of each respondent that opted in ( $N$  of approximately 5000). Second, we randomly assigned each of these unique identifiers to one of the eight researcher name conditions listed in Table 1. Finally, we created separate MTurk accounts under each researcher name and deployed the same survey within each account. Subjects were assigned a "qualification" within the MTurk interface,

<sup>9</sup>The results for this are presented in Section D of the Online Appendix.

<sup>10</sup>The question asked about the number of tasks the respondent had previously completed on Mechanical Turk.

according to their assigned condition. Each survey was set such that only MTurk workers with the correct qualification could see the survey (and thus the username associated with it).<sup>11</sup> This meant that each potential respondent could see only one survey from their assigned researcher, and could then choose whether or not to take that survey. In summary, we posted an initial survey where we collected MTurk IDs, randomly assigned these workers to one of eight conditions where we varied the researcher name, and then *only respondents in that condition could view that HIT*.<sup>12</sup>

Within the survey, respondents answered a series of questions about social and political attitudes. We drew questions from Pew, Gallup, and the American National Election Survey, specifically asking about issues for which racial and gender cues may prompt different responses.<sup>13</sup> We chose to ask questions about race and gender, as these are two of the main areas where prior research has demonstrated that interviewer attributes can affect subject behavior. Moreover, this is the information conveyed most prominently by researchers through their names in online surveys. After all subjects had completed all study-related activities, respondents were debriefed about the nature and purpose of the study.

## RESULTS

Our design allows us to test whether researcher identity shapes the sample of respondents that agree to take the survey. We find little evidence of such an effect.<sup>14</sup> We find substantively small differences in the number of people who take the different surveys, and no difference in respondents' backgrounds on a range of personal characteristics. We also do not find differences in survey completion rates across name; all rates were extremely high (above 97%). Therefore, we are not concerned about inducing selection bias by analyzing the set of completed surveys. We turn next to the content of survey responses.

Our analyses fail to reject the null hypothesis that there is no difference in how respondents answer questions when assigned to a putatively black or female researcher relative to a white or male one. We estimate all of our treatment effects using linear regression models, regressing outcome on the indicator of treatment.

<sup>11</sup>In practice, Mechanical Turk functions were done through R scripts using the MTurkR package to access the Mechanical Turk API (Leeper, 2015, 2013). This allowed us to post tasks in small batches (of 9 at a time) so as to avoid having the tasks posted to online MTurk discussion boards where workers share lucrative HIT opportunities (this could have exposed our experimental design). We posted these small batches at short, regular intervals (each HIT expired and was re-posted every 15 min for several days) to ensure that the tasks were continuously available to potential workers across all experimental conditions. This approach seems to have worked: regular scans of major MTurk discussion boards did not reveal any postings about our HITs.

<sup>12</sup>"HITs" or Human Intelligence Tasks are the name MTurk gives any individual unit of work posted on the site. In this case, a HIT included a link to take our survey for some pre-specified payment amount.

<sup>13</sup>The full text of the outcome questions is presented in Section I of the Online Appendix.

<sup>14</sup>We explore the selection process in more detail in the Online Appendix.

Robust standard errors are estimated using a bootstrapping procedure. Following our pre-analysis plan, our rejection levels for accepting that the effects differ from zero are calibrated to yield an expected number of false discoveries of  $\alpha = 0.05$ , adjusting for multiple testing using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995).<sup>15</sup> This adjustment is important since we dramatically increase the chances of a false positive finding by testing for multiple outcomes (Benjamini and Hochberg, 1995).<sup>16</sup> To avoid the appearance of “fishing” for significant  $p$ -values across many outcomes, we cannot simply follow a rule of rejecting any null hypothesis when  $p < 0.05$ . We focus on estimating only the average treatment effects of the researcher race and gender treatments, and, consistent with our pre-analysis plan, only investigate possible treatment effect heterogeneity as exploratory rather than confirmatory results.<sup>17</sup>

Our first set of outcome questions examines whether assignment to a putatively female/black (relative to male/white) investigator changes reported affect towards, or support for policies meant to help, women/blacks.<sup>18</sup> For the race dimension of treatment, we estimate treatment effects on three distinct outcomes: expressed racial resentment (as measured by the 0–1 scale developed by Kinder and Sanders, 1996), willingness to vote for a black president, and support for social service spending. On gender, we examine respondents’ beliefs regarding the role of women in society, willingness to vote for a woman presidential candidate, and the same social service spending outcome. In selecting our first two outcome questions, we sought questions that were both commonly used in online surveys but also directly related to each of our treatments. The social spending measure was included as a facially non-racial measure that could still have racial or gendered overtones. This allowed us to test whether respondents would think of social spending as disproportionately benefiting minorities and women, and so potentially answer in either raced or gendered ways depending on the putative race or gender of the researcher.

<sup>15</sup>For the Benjamini–Hochberg procedure, we order our  $m$   $p$ -values from smallest to largest  $P_{(1)}, \dots, P_{(m)}$  and find the largest  $k$  such that  $P_{(k)} \leq \alpha \times \frac{k}{m}$ . This ensures that our false discovery rate, that is the expected share of rejected nulls that are “false positives,” is controlled at  $\alpha = 0.05$ . Note that under this procedure, we would not reject any nulls if all  $p$ -values are  $> 0.05$  and reject all nulls if all  $p$ -values are  $< 0.05$ .

<sup>16</sup>While Benjamini and Hochberg (1995) analyze the case where hypotheses are independent, Benjamini and Yekutieli (2001) show that the procedure also properly controls the false-discovery rate under positive dependence between hypotheses. This is likely to be the case under our set of tests as each question can be seen as measuring different elements of an individual’s latent affect towards a group. Moreover, simulation studies by Groppe et al. (2011) show good performance of the Benjamini–Hochberg method even under violations of independence.

<sup>17</sup>For a discussion of potential treatment effect heterogeneity by race/gender, see Section G of the Online Appendix.

<sup>18</sup>In Sections C– E of the Online Appendix, we also report results for selection into the survey itself, survey completion, and attention check passage rates, finding no substantive differences across the treatment conditions.

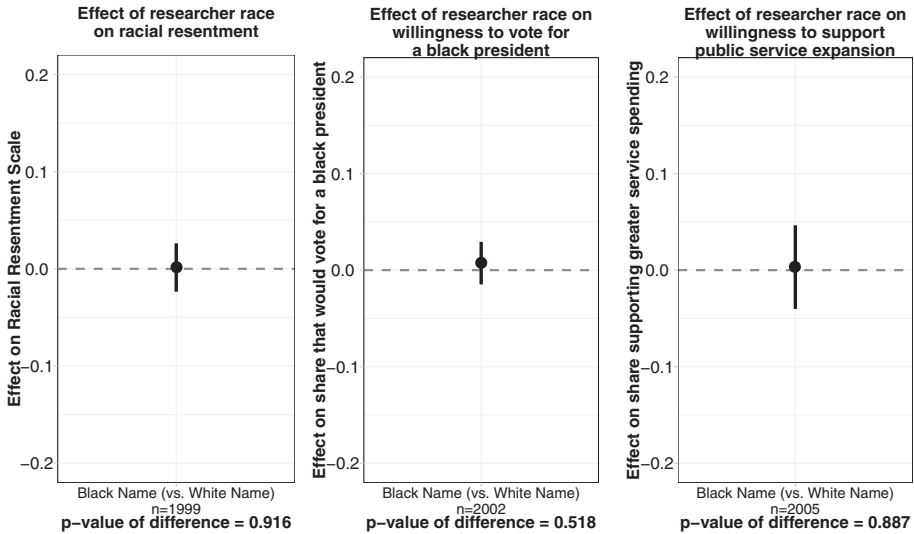


Figure 1

### Difference in Policy/Attitude Outcomes for Researcher Race Treatment

Lines denote 95% multiple comparison adjusted confidence intervals (Benjamini and Yekutieli, 2005).

We designed our experiment to target a sample size of 2000 total respondents.<sup>19</sup> For the race treatment, we find no evidence that black versus white researcher names yield different responses on the outcome questions. Figure 1 plots the expected difference in outcomes for each of these three questions for respondents assigned to a hypothetical black researcher name relative to respondents assigned to a hypothetical white researcher name. For all three outcomes, the difference in outcomes between the two treatment groups is not statistically significant at  $\alpha = 0.05$ . We fail to reject the null of no effect for all outcomes at the  $\alpha = 0.05$  level.

For the gender treatment, when we adjust for multiple comparisons we fail to reject the null hypothesis that there is no difference between putatively male or female researchers. Figure 2 plots the difference in expected values for each of the the outcomes between the female researcher and male researcher treatment conditions. While we fail to reject the null, we should note that for all outcomes, respondents under the female researcher treatment condition were about 2–4 percentage points more likely to express affective/policy support for women. The individual  $p$ -value for the null of no effect on the gender equality outcome question fell just below the commonly used threshold of 0.05. The  $p$ -values for the null for the other two outcomes, however, fall just above the typically used threshold. Under

<sup>19</sup>Our final sample consists of 2006 unique respondents that we could confirm had completed the overall survey. We omit responses from one respondent who requested that their responses not be used after the debriefing process.

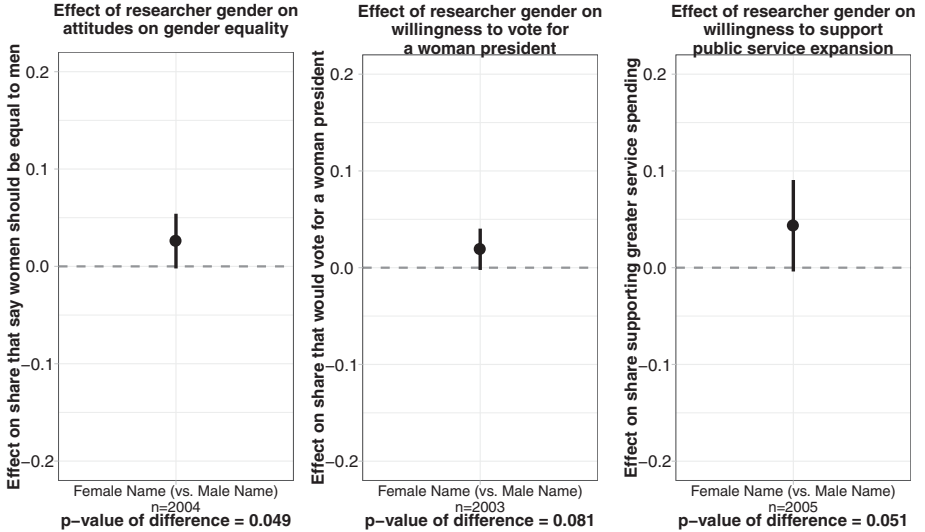


Figure 2

### Difference in Policy/Attitude Outcomes for Researcher Gender Treatment

Lines denote 95% multiple comparison adjusted confidence intervals (Benjamini and Yekutieli, 2005).

our pre-registered design, using the Benjamini–Hochberg correction for multiple testing, we fail to reject the null for all three outcomes.<sup>20</sup> We cannot conclude that assignment to a putatively female researcher name significantly increased the likelihood that respondents would exhibit more woman-friendly attitudes on gender-related questions.

Despite our failure to reject the null, we note that the point estimates for the direction of the effect are consistent with our original hypothesis. In general, respondents assigned to a putatively female investigator were, in-sample, more likely to express beliefs that were more supportive of women's equality. Given this, how concerned should researchers be about these estimates? Power calculations for our design suggest a relatively small upper bound for any “true” effect. For a study of our sample size, and accounting for the multiple comparison adjustment, we conclude that it is unlikely that we would have failed to reject had the true effect of any one of these outcomes been greater than 5 percentage points.<sup>21</sup> While it is not possible to “affirm” a null hypothesis, the high power of our study is such that our null finding implies any real effect is likely to be bounded close to 0.

<sup>20</sup>This is because the threshold level for rejecting the null of no effect on only the gender equality outcome falls to  $\alpha = 0.05 \times \frac{1}{3} = 0.016$ —below the  $p$ -values that we observe.

<sup>21</sup>For a more detailed discussion of the power calculations, see Section F of the Online Appendix.



## DISCUSSION AND CONCLUSION

In this paper, we demonstrate that researchers using online survey platforms such as Amazon's Mechanical Turk generally need not be concerned that information conveyed through their name in either the advertisement for the HIT or the informed consent page will subsequently affect their results. Our study is designed to address both elements of investigator bias: inferences about the purposes of the study and comfort with the investigator, either of which we might expect to affect the willingness of respondents to take the survey in the first place, overall effort, and the types of answers given. We fail to reject the null hypothesis that researchers' race or gender (cued through names) have no effect on respondents' survey behaviors. While our evidence suggests that there might be a small "true effect" of researcher gender, our power calculations demonstrate that this effect, if any, is quite small and likely not substantively meaningful for most researchers.

There are at least two plausible explanations for why the results of this paper diverge from the substantively meaningful effects found in research on other survey platforms. First, it could be the case that either the strength or substance of the treatment differs between online survey platforms and other modes of conducting surveys (such as in-person or telephone). That is, interacting in person with a black/white or male/female researcher might have a stronger effect on respondent behavior than simply reading their names. Substantively, this means that even if respondents do notice the putatively black/white or male/female name assigned to the researcher through treatment, the act of reading this name is simply not enough to change their subsequent survey responses.

Second, it could be the case that respondents in online survey platforms are less likely to take treatment.<sup>22</sup> If this were the case, it could be in part driven by the fact that our respondents were recruited via Amazon's Mechanical Turk, where the financial incentives for respondents to complete tasks as quickly as possible might lead them to quickly skim through the consent page.<sup>23</sup> This means they would be less likely to notice the researcher name and thus less likely to respond to it. Even if respondents were prone to bias, it could be masked by the fact that few respondents actually read the names in the first place. The present study is unable to adjudicate between these two potential explanations.

For researchers conducting studies on MTurk and similar online platforms, this distinction will not matter. Nevertheless, the two different mechanisms have important implications both for the external validity of the present study as well as

<sup>22</sup>This explanation would be consistent with the null finding presented in this paper on the effect of researcher name on likelihood of selecting into the survey, though this evidence is not sufficient to rule out the first explanation.

<sup>23</sup>We note, though, that MTurk workers' financial incentives could operate in either direction. The existence of services like Turkopticon, used to keep track of individual requester accounts and their payment practices, suggests that Turkers might be even more motivated than other survey takers to notice researcher names.

further research on the attributes of online survey pools. In particular, researchers should be cautious in applying the results of this study when either (1) they provide more information about themselves than simply their name in the advertisement for the survey and on the informed consent page (that is, they have a stronger treatment), or (2) respondents in their sample pay more attention throughout all stages of the survey than MTurk respondents (i.e., there is higher treatment uptake). In our experience, the first point is unlikely to occur across different survey platforms since a few platforms provide more researcher information than MTurk. However, whether and how much attention varies across different survey platforms and how this substantively affects results is an open question and interesting area of further research.

## SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit <https://doi.org/10.1017/XPS.2017.25>

## REFERENCES

- Adida, Claire L., Karen E. Ferree, Daniel N. Posner, and Amanda Lea Robinson. 2016. "Who's Asking? Interviewer Coethnicity Effects in African Survey Data." *Comparative Political Studies* 49(12): 1630–1660.
- Benjamini, Yoav and Daniel Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Annals of Statistics* 29(4): 1165–1188.
- Benjamini, Yoav and Daniel Yekutieli. 2005. "False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters." *Journal of the American Statistical Association* 100(469): 71–81.
- Benjamini, Yoav and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289–300.
- Berinsky, Adam J., Michele F. Margolis, and Michael W. Sances. 2014. "Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys." *American Journal of Political Science* 58(3): 739–753.
- Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. 2003. "The Advent of Internet Surveys for Political Research: A Comparison of Telephone and Internet Samples." *Political Analysis* 11(1): 1–22.
- Bertrand, Marianne and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94(4): 991–1013.
- Bush, Sarah and Lauren Prather. 2017. "How Electronic Devices in Face-to-Face Interviews Change Survey Behavior: Evidence From a Developing Country." ([http://www.laurenprather.org/uploads/2/5/2/3/25239175/bush\\_prather\\_electronic\\_devices\\_in\\_survey\\_interviews.pdf](http://www.laurenprather.org/uploads/2/5/2/3/25239175/bush_prather_electronic_devices_in_survey_interviews.pdf)), accessed December 29, 2017.
- Butler, Daniel M. and David E. Broockman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55(3): 463–477.

- Butler, Daniel M. and Jonathan Homola. 2017. "An Empirical Justification for the Use of Racially Distinctive Names to Signal Race in Experiments." *Political Analysis* 25(1): 122–130.
- Chandler, Jesse, Pam Mueller, and Gabriele Paolacci. 2014. "Nonnaïveté Among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers." *Behavior Research Methods* 46(1): 112–130.
- Charnysh, Volha, Christopher Lucas, and Prerna Singh. 2014. "The Ties That Bind: National Identity Salience and Pro-Social Behavior Toward the Ethnic Other." *Comparative Political Studies* 48(3): 267–300.
- 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.
- Cotter, Patrick R., Jeffrey Cohen, and Philip B. Coulter. 1982. "Race-of-Interviewer Effects in Telephone Interviews." *Public Opinion Quarterly* 46(2): 278–284.
- Davis, Darren W. 1997. "The Direction of Race of Interviewer Effects Among African-Americans: Donning the Black Mask." *American Journal of Political Science* 41(1): 309–322.
- Davis, Darren W. and Brian D. Silver. 2003. "Stereotype Threat and Race of Interviewer Effects in a Survey on Political Knowledge." *American Journal of Political Science* 47(1): 33–45.
- Edelman, Benjamin, Michael Luca, and Dan Svirsky. 2017. "Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment." *American Economic Journal: Applied Economics* 9(2): 1–22.
- Einstein, Katherine and David M. Glick. 2017. "Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats and Access to Public Housing." *American Journal of Political Science* 61(1): 100–116.
- Fryer, Jr., Roland G., and Steven J. Levitt. 2004. "The Causes and Consequences of Distinctively Black Names." *The Quarterly Journal of Economics* 119(3): 767–806.
- Grewal, Ini and Jane Ritchie. 2006. "Ethnic and Language Matching of the Researcher and the Research Group During Design, Fieldwork and Analysis." In *Health and Social Research in Multiethnic Societies*, ed. James Y. Nazroo. Oxon: Routledge, 65–81.
- Groppe, David M., Thomas P. Urbach, and Marta Kutas. 2011. "Mass Univariate Analysis of Event-Related Brain Potentials/Fields II: Simulation Studies." *Psychophysiology* 48(12): 1726–1737.
- Hatchett, Shirley and Howard Schuman. 1975. "White Respondents and Race-of-Interviewer Effects." *The Public Opinion Quarterly* 39(4): 523–528.
- Healy, Andrew and Gabriel S. Lenz. 2014. "Substituting the End for the Whole: Why Voters Respond Primarily to the Election-Year Economy." *American Journal of Political Science* 58(1): 31–47.
- Huddy, Leonie, Joshua Billig, John Bracciodieta, Patrick J. Moynihan, and Patricia Pugliani. 1997. "The Effect of Interviewer Gender on the Survey Response." *Political Behavior* 19(3): 197–220.
- Huff, Connor and Dustin Tingley. 2015. "Who Are These People? Evaluating the Demographic Characteristics and Political Preferences of MTurk Survey Respondents." *Research & Politics* 2(3): 1–12.
- Kinder, Donald R. and Lynn M. Sanders. 1996. *Divided by Color: Racial Politics and Democratic Ideals*. Chicago: University of Chicago Press.
- Kriner, Douglas L. and Francis X. Shen. 2013. "Reassessing American Casualty Sensitivity: The Mediating Influence of Inequality." *Journal of Conflict Resolution* 58(7): 1174–1201.

- Krupnikov, Yanna and Adam Seth Levine. 2014. "Cross-Sample Comparisons and External Validity." *Journal of Experimental Political Science* 1(1): 59–80.
- Leeper, Thomas. 2013. "Crowdsourcing with R and the MTurk API." *The Political Methodologist* 20: 2–7.
- Leeper, Thomas J. 2015. *MTurkR: Access to Amazon Mechanical Turk Requester API via R*. R package version 0.6.5.1. (<https://cran.r-project.org/web/packages/MTurkR/MTurkR.pdf>), accessed December 29, 2017.
- Leeper, Thomas and Emily Thorson. 2015. "Minimal Sponsorship-Induced Bias in Web Survey Data." (<https://s3.us-east-2.amazonaws.com/tjl-sharing/assets/SurveySponsorship.pdf>), accessed December 29, 2017.
- Levy, Kevin E., Jeremy Freese, and James N. Druckman. 2016. "The Demographic and Political Composition of Mechanical Turk Samples." *SAGE Open* 6(1): 1–17.
- Mullinix, Kevin J., Thomas J. Leeper, James N. Druckman, and Jeremy Freese. 2015. "The Generalizability of Survey Experiments." *Journal of Experimental Political Science* 2(2): 109–138.
- Mummolo, Jonathan and Erik Peterson. 2017. "Demand Effects in Survey Experiments: An Empirical Assessment." ([https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2956147](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2956147)), accessed December 29, 2017.
- Reese, Stephen D., Wayne A. Danielson, Pamela J. Shoemaker, Tsan-Kuo Chang, and Hwei-Ling Hsu. 1986. "Ethnicity-of-Interviewer Effects Among Mexican-Americans and Anglos." *Public Opinion Quarterly* 50(4): 563–572.
- Schaeffer, Nora Cate, Jennifer Dykema, and Douglas W. Maynard. 2010. "Interviewers and Interviewing." In *Handbook of Survey Research*, eds. Peter V. Marsden and James D. Wright. Bingley: Emerald, 437–470.
- Survey Research Center. 2010. *Guidelines for Best Practice in Cross-Cultural Surveys*. Ann Arbor, MI: Survey Research Center, Institute for Social Research, University of Michigan. (<http://www.ccsr.isr.umich.edu/>), accessed December 29, 2017.
- Tomz, Michael and Jessica Weeks. 2013. "Public Opinion and the Democratic Peace." *American Political Science Review* 107(4): 849–865.
- Wallace, Geoffrey P. R. 2013. "International Law and Public Attitudes Toward Torture: An Experimental Study." *International Organization* 67(1): 105–140.
- White, Ariel R., Noah L. Nathan and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109(1): 129–142.
- Word, David L., Charles D. Coleman, Robert Nunziata and Robert Kominski. 2008. "Demographic Aspects of Surnames from Census 2000." (<https://www2.census.gov/topics/genealogy/2000surnames/surnames.pdf>), accessed December 29, 2017.