# Investigator Characteristics and Respondent Behavior in Online Surveys [*]

Connor Huff[†], Dominika Kruszewska[‡], Christopher Lucas[§],
Anton Strezhnev,[¶] Ariel White[‖]

This Draft: July 17, 2015

## 1   Introduction

A long strain of research in political science has shown how the responses of interviewees in face-to-face and telephone surveys can vary depending on the race or gender of the interviewer (Davis, 1997; Davis & Silver, 2003; Hatchett & Schuman, 1975; Cotter et al., 1982; Reese et al., 1986; Huddy et al., 1997). This variation means that the inferences researchers draw and the replicability of the study can depend on who runs the study. Regardless of whether the relationship between characteristics of the interviewer and the observed variation in responses is perceived to be a benefit[1] or a threat,[2] it is useful for researchers to be cognizant of the circumstances in which we might expect this relationship to be most likely to occur. For example, it is often suggested that responses in online surveys are less likely to be affected by attributes of the researcher than in other survey methods.[3]

---

[*]Please do not circulate without permission of the authors.

[†]Department of Government, Harvard University, cdezzanihuff@fas.harvard.edu

[‡]Department of Government, Harvard University, dkruszewska@fas.harvard.edu

[§]Department of Government, Harvard University, clucas@fas.harvard.edu

[¶]Department of Government, Harvard University, astrezhnev@fas.harvard.edu

[‖]Department of Government, Harvard University, arwhite@fas.harvard.edu

[1]This is most likely to occur when respondents are more comfortable with an interviewer and thus more likely to express their true preferences.

[2]The primary concern is that this challenges the external validity of the findings.

[3]Indeed, Dillman & Christian (2005) write that social desirability is "more likely to influence answers in interview surveys than in those that use self-administration [like internet-based surveys] because of the pres-

In this paper we explore how attributes of the researcher affect responses in online surveys. In particular, we use a survey experiment that explicitly manipulates the race and gender cued by the researcher name on the informed consent page. The informed consent page is generally required by the Institutional Review Board (IRB) of research universities to be displayed at the start of every internet-based survey. Manipulating the researcher name allows us to test how information conveyed about the race and gender of the researcher through the informed consent page affects survey responses. We focus on gender and race because these two factors can be clearly conveyed through names (Bertrand & Mullainathan, 2004; Milkman et al., 2012), and have been central to the existing literature on surveys and identity. The experiment is a 2x2 factorial design, where the first factor is the putative gender of the investigator (male or female) and the second is the investigator's putative race (white or black).

In this research registration, we present our hypotheses and expectations in Section 2. Section 3 lays out the experimental design (as a supplement to this section, Appendix A contains the survey in full). Section 4 clarifies the tests we'll conduct to evaluate these hypotheses. Section 5 concludes.

## 2 Hypotheses

There are two main ways we might theoretically expect the putative name of the researcher to affect responses. First, researcher identity could shape the responses given to questions about race or gender, as in the existing race-of-interviewer literature. Respondents would be more likely to give answers suggesting positive feelings toward women or black people, or support for equal status for these groups, when the researcher is assumed to belong to the

---

ence of an interviewer." Similarly, in a lab experiment, Chang & Krosnick (2010) find less social desirability response bias in the computer mode than in the intercom mode, pointing to this increase in honest answers as one of the potential advantages of computer survey administration, while Kreuter et al. (2008) show that web administration increases reporting of sensitive information and its accuracy. This strain of research suggests that for researchers concerned about social desirability bias affecting responses, online-surveys might present a promising fora to reach a large number of respondents while limiting social desirability bias.

groups in question.

> **$H_1$** : *Assignment to an investigator name that is commonly perceived to be female/black will increase reported support for policies that provide for and protect the rights of women/blacks, and decreases responses that indicate prejudice against these groups.*

The second way we might theoretically expect the putative name of the researcher to affect responses is through varying levels of effort. This variation in effort is most likely to occur on platforms like Amazon's Mechanical Turk, where respondents are paid based on the number of HITs they complete. In these scenarios, respondents have incentives to attempt to complete HITs as quickly as possible. Thus, if they think that individuals of a certain race/gender are more/less likely to punish inattentiveness then they will be more/less likely to exert more effort. For example, according to gender stereotypes, women are expected to show greater warmth and behave more communally than men, whereas aggressiveness and competitiveness is perceived as more congruent with the male role.[4] Thus, respondents might be more likely to think that they will be punished by males than females.

Additionally, the perception of the status of the researcher's group and thus the researcher's authority might influence effort exerted by the subjects. Similar to a "lab coat" effect, where clothes cue competence and authority impacting subject behavior, the identity of the researcher might signal domination and expertise, prompting the participants to try harder to perform well and be attentive when faced with a researcher perceived to belong to a higher status or more dominant group.[5]

This has important implications for political science researchers, since the ability to detect treatment effects hinges on respondents actually taking treatment. This means that regardless of the topic of the survey we might expect to observe varying treatment effects

---

[4]See for example Carli (2001).

[5]For a discussion of the role of gender in exterting social influence, see Carli (2001). For a an overview of studies on gender and race differences in job authority, see Smith (2002).

depending on whether the researcher is white/black or male/female. This has potentially profound implications for political science research as it could contribute to some publication biases prevalent throughout social science research (people are more likely to cite and co-author with males, etc.) if researchers in socially dominant groups are more likely to find treatment effects, particularly in experimental designs which rely on subtle differences in wording. We derive our second hypothesis directly from this logic.

>*$H_2$ : Attention and effort will be greatest among subjects assigned to a putatively white, male investigator.*

## 3 Experimental Survey Design

We plan to test the above hypotheses using a survey experiment fielded online. Respondents will be recruited online via Amazon's Mechanical Turk (MTurk) platform. Many experimental researchers have begun using MTurk for low-cost recruitment of subjects, making it a highly relevant subject pool for investigating the effects of researcher identity within an online context. All respondents are exposed to a standard IRB consent form on the first page of the survey. On this page, the name of the researcher is randomly selected from a list of 40 possible names meant to cue racial and gender identity. The researcher's gender is presented as either male or female and race is presented as either black or white. To generate the names associated with each of these manipulations, we combined three commonly used lists of racially distinct first and last names. First names were drawn from a combination of lists found in Bertrand & Mullainathan (2004) and Fryer, Jr. & Levitt (2004), while last names were drawn from lists in Word et al. (2008) and Bertrand & Mullainathan (2004). We crossed the lists of first and last names to produce many possible combinations[6] and randomly drew

---

[6]We omitted a few randomly-generated names that already belonged to celebrities, such as Jermaine Jackson.

ten names for each of the four manipulation categories (black women, white women, black men, and white men). The full list of names used in this experiment is presented in Table 1.

| Black Men | White Men | Black Women | White Women |
|---|---|---|---|
| Tyrone Booker | Cody Novak | Tanisha Dorsey | Molly Koch |
| Rasheed Washington | Neil Kelly | Deja Gaines | Molly Baker |
| Reginald Jones | Connor Sullivan | Deja Banks | Carrie Haas |
| Darnell Washington | Greg Sullivan | Latoya Dorsey | Molly Sullivan |
| Darnell Banks | Brett Schwartz | Aisha Rivers | Abigail Koch |
| Kareem Dorsey | Brad Walsh | Keisha Williams | Kristen Murphy |
| Rasheed Jackson | Connor McCarthy | Tanisha Rivers | Carrie McCarthy |
| Darnell Booker | Matthew Schmitt | Keisha Joseph | Molly Murphy |
| Rasheed Dorsey | Greg Murphy | Shanice Williams | Abigail Ryan |
| Tyrone Gaines | Todd McCarthy | Aisha Mosley | Carrie Ryan |

Table 1: Names used for each of the four investigator name manipulations, based on lists from Bertrand & Mullainathan (2004), Fryer, Jr. & Levitt (2004), Word et al. (2008)

After the consent page manipulation, respondents are asked a series of questions about social and political attitudes. We draw questions from Pew, Gallup, and the American National Election Survey specifically asking about issues for which racial and gender cues may prompt different responses. The full text of the outcome questions is presented in Appendix A. We also explore the extent to which respondents are paying attention and are willing to put in effort using attention checks and open text responses. The general structure of the attention checks used in the experiment is taken from Berinsky et al. (2014). Respondents are also asked to complete a randomly assigned writing task, either on their attitudes towards a female president or on a time in their life when they were affected by politics. The latter prompt is sufficiently general that variation in response depth will capture respondent's general "effort" levels rather than attitudes towards a particular issue. In order to obscure the general purpose of the survey, we randomly permute some of the demographic questions with the outcome questions. However, to avoid priming party affiliation, gender or race, we leave the party ID, gender, and race questions for the end of the survey.

# 4    Statistical Tests

For Hypothesis 1, we estimate two separate treatment effects. The first is the effect of assignment to a putatively female name on the probability that a respondent indicates that they believe that women should have an equal role in the workforce. The second is the effect of assignment to a putatively black name on the respondent's racial resentment scale. We expect that the effect for the former will be positive while the latter will be negative. For Hypothesis 2, we will estimate the effect of assignment to a putatively white and male name on the probability that a respondent correctly completes both of the attention check assignments. We expect this effect estimate to be positive. For estimation, we will fit a linear probability model of the outcome on treatment and compute standard errors via a nonparametric bootstrapping procedure. While not needed for identification, we will include respondent-level covariates (e.g. gender, income, education) in the regression model in order to increase the efficiency of our estimator.

Because respondents have the option to stop taking the survey after treatment is assigned, there is concern that an analysis conditional on survey completion will be biased for the average treatment effect if treatment also affects the probability that a respondent will drop out. To obtain unbiased treatment effect estimates in this situation, we adopt an estimation strategy similar to that of Rotnitzky & Robins (1995) and weight each respondent observation in the outcome regression by its estimated probability of not dropping out of the sample. We estimate this probability via a logistic regression of completion on treatment using the entire set of respondents (those that both completed and did not complete the survey).

Our rejection levels for two-sided hypothesis tests of whether the average treatment effects differ from zero are calibrated to correct for problems of multiple testing. We are willing to tolerate an overall Type I error rate of $\alpha = .05$. With three main hypothesis tests, we could obtain a conservative rejection threshold for each individual hypothesis test of $.05/3 = .017$ using the Bonferroni correction. This controls the Familywise Type I Error Rate and guarantees that the probability of any single erroneous rejection in the set of

tests is less than or equal to .05. However, this approach sacrifices a significant amount of power. A less conservative but more powerful approach is to set a rejection threshold to control the False Discovery Rate (FDR). We use the Benjamini-Hochberg procedure to set a rejection level for the hypothesis tests (Benjamini & Hochberg, 1995). This entails a two-step procedure where we order the 3 p-values of the individual hypothesis tests from smallest to largest, $p_{(1)}, ..., p_{(3)}$ and then set our rejection level to $p_{(k)}$, where $k$ is the largest value of $i$ that satisfies $p_{(i)} \leq \frac{i}{3}\alpha$. This procedure controls the expected share of false hypothesis rejections out of the total number of rejections to be no greater than .05.

We do not specify any ex-ante interactions of the treatment effects with baseline covariates. However, because the mechanism through which any treatment effects operate are of significant interest, we will conduct exploratory analyses of potential treatment effect heterogeneity by estimating models with interactions between treatment and respondent identity variables. Among other interactions, we are interested in seeing whether any average treatment effect is primarily driven by behavior changes among men (in the case of the gender treatment) and white respondents (in the case of the race treatment). We will attempt to replicate any promising results from these exploratory analyses in a follow-up experiment that explicitly registers interactive hypotheses prior to the experiment.

# 5   Next Steps

In this preregistration plan, we outline an experiment that tests whether or not investigator characteristics have an effect on subjects responses and subject effort. This design permits direct tests of these hypotheses. In addition to these primary hypothesis tests, we hope to conduct an exploratory analyses of heterogeneous treatment effects, which will serve as the basis for a second experiment to test the mechanisms that we hope to identify in the experiment laid out in this plan. The second experiment will be preregistered separately, given that its design depends on the results of the experiment outlined here.

# A   Appendix: Survey Text

## A.1   Informed Consent

Thank you for participating in this survey. Please take time to answer questions honestly and thoroughly. Your responses are essential to our research.

This research is being conducted under the supervision of BLACK/WHITE x MALE/FEMALE NAME at Harvard University. All of the information that we obtain from your session will be anonymous. We do not ask you for your name. Your name or identifying information will not be used in any reports of the research. There will be no direct benefit to you from participation in this study other than the agreed-upon financial compensation. We hope, however, that the research will benefit society by improving our understanding of the factors that influence people's decision making.

If you want to receive the findings of this study, you may contact NAME HERE (GENERIC EMAIL). Complete contact information is as follows.

**NAME HERE**

Department of Government

Harvard University

Cambridge, MA 02138

GENERIC EMAIL HERE

If you have questions about your rights or about research-related harm, or if your questions, concerns, suggestions, or complaints are not being addressed by the researchers above, please contact:

Director of IRB Operations

Harvard University Committee on the Use of Human Subjects in Research

1414 Massachusetts Avenue, Second Floor

Cambridge, MA 02138

Phone: 617-496-5593

jjaeger@fas.harvard.edu


The nature and purpose of this study have been satisfactorily explained to me and I (participant) agree to become a participant in the study described above. I understand that I am free to discontinue participation at any time if I so choose. Answering yes continues the survey, answering no ends the survey.

Yes

No


## A.2 Outcome Questions [Question order randomized]

- Do you agree strongly, agree somewhat, neither agree nor disagree, disagree,somewhat, or disagree strongly with these statements?

  - Over the past few years, blacks have gotten less than they deserve.

  - Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors.

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

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


- Some people feel that women should have an equal role with men in running business,

industry and government. Others feel that women's place is in the home. Where would you place yourself on this scale or haven't you thought much about this?

Equal role

Womens place is in the home

Havent thought much about this

- Between now and the 2016 Presidential Election, there will be discussion about the qualifications of presidential candidates - their education, age, race, religion, and so on. If your party nominated a generally well-qualified person for president who happened to be _____, would you vote for that person?

Black ("Yes, would" or "No, would not")

A woman ("Yes, would" or "No, would not")

Catholic ("Yes, would" or "No, would not")

Hispanic ("Yes, would" or "No, would not")

Jewish ("Yes, would" or "No, would not")

Mormon ("Yes, would" or "No, would not")

Gay or lesbian ("Yes, would" or "No, would not")

Muslim ("Yes, would" or "No, would not")

An atheist ("Yes, would" or "No, would not")

- Some people think the government should provide fewer services, even in areas such as health and education, in order to reduce spending. Other people feel that it is important for the government to provide many more services even if it means an increase in spending. Which do you prefer?

Cut services/spending

More services/spending

# Demographic Questions [Order permuted with outcome questions]

- What is the highest level of education you have completed?

  Less than High School

  High School / GED

  Some College

  2-year College Degree

  4-year College Degree

  Masters Degree

  Doctoral Degree

  Professional Degree (JD, MD)


- What is your yearly household income, putting together the income of all the members of your household?

  Less than 30,000

  30,000  39,999

  40,000  49,999

  50,000  59,999

  60,000  69,999

  70,000  79,999

  80,000  89,999

  90,000  99,999

  100,000 or more

## Attention Checks [permuted with the outcome questions]

- When a big news story breaks people often go online to get up-to-the-minute details on what is going on. We want to know which websites people trust to get this information. We also want to know if people are paying attention to this survey. To show that you've read this much, please ignore the question and select Reuters website and Huffington Post as your two answers.

  When there is a big news story, which is the one news website you would visit first? (Please only choose one)

  | | | |
  | :---: | :---: | :---: |
  | New York Times website | The Drudge Report | The Associated Press (AP) website |
  | Huffington Post | Google News | Reuters website |
  | Washington Post website | ABC News website | National Public Radio (NPR) website |
  | CNN.com | CBS News website | USA Today website |
  | FoxNews.com | NBC News website | New York Post Online |
  | MSNBC.com | Yahoo! News | None of these websites |

- We are very interested to know what political issues people think are the most relevant today. People often have different attitudes about what issues the United States government should focus on addressing and we would like to understand more about this public debate. We also want to know if people are paying attention to this question. To show that you've read this much, please ignore the question and select Energy and Global trade as your two answers.

  Which of the following issues do you think should be the highest priority for President Obama and Congress in 2015? (Please only choose one)

|            |                 |                      |
|------------|-----------------|----------------------|
| Terrorism  | Reducing crime  | Energy               |
| Economy    | Poor and needy  | Influence of lobbyists |
| Jobs       | Military        | Transportation       |
| Education  | Immigration     | Money in politics    |
| Social Security | Environment | Scientific research  |
| Budget deficit | Race relations | Global warming      |
| Health care costs | Moral breakdown | Global trade     |
| Medicare   | Tax reform      | None of these issues |

## Open Response Questions

[Randomly assign respondents to one of the two questions below]

- Please write a few sentences about what you think about the United States potentially having a female president.

- Please write a few sentences about a time that politics affected your life.

## Demographic Questions [placed at end of survey]

- Generally speaking, do you consider yourself to be a(n):

  Democrat

  Republican

  Independent


- → [if independent] As of today do you lean more to the Republican Party or more to the Democratic Party?

  Republican

Democrat

- In talking to people about elections, we often find that a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have time. How about you – did you vote in the last presidential election in 2012?

  I don't remember.

  No, I did not vote.

  Yes, I voted.

- → [if yes] Who did you vote for in the last presidential election?

  Mitt Romney

  Barack Obama

  Other (specify)

- What is your ethnicity? Select all that apply.

  Black

  White

  Hispanic

  Asian

  Native American

  Other (specify)

- What is your gender?

  Male

  Female

Other (please specify)

- What do you think is the purpose of this research?

  [Open Response]

- While taking this survey, did you engage in any of the following behaviors?

  Use your cell phone

  Browse the internet

  Talk with another person

  Watch TV

  Listen to music

## Debriefing (sent to all subjects by email AFTER all responses are collected)

The purpose of this study was to learn about how researcher identity influences subjects' responses. In order to learn about this, we needed to manipulate the identity of the researcher. Though the contact information you were given was in fact real, the name of the researcher was not. This study was conducted by Connor Huff, Dominika Kruszewska, Christopher Lucas, Anton Strezhnev, and Ariel White, all at Harvard University. If you have any questions about this study, please contact them at the following location.

Connor Huff, Dominika Kruszewska, Christopher Lucas, Anton Strezhnev, Ariel White

Department of Government

Harvard University

Cambridge, MA 02138

GENERIC EMAIL HERE

# References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.

Berinsky, A. J., Margolis, M. F., & Sances, M. W. (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. doi: 10.1111/ajps.12081

Bertrand, M., & Mullainathan, S. (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. doi: 10.1257/0002828042002561

Carli, L. L. (2001). Gender and social influence. *Journal of Social Issues*, *57*(4), 725–741.

Chang, L., & Krosnick, J. A. (2010). Comparing oral interviewing with self-administered computerized questionnairesan experiment. *Public Opinion Quarterly*, nfp090.

Cotter, P. R., Cohen, J., & Coulter, P. B. (1982). Race-of-Interviewer Effects in Telephone Interviews. *Public Opinion Quarterly*, *46*(2), 278–284. doi: 10.1086/268719

Davis, D. 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. doi: 10.2307/2111718

Davis, D. W., & Silver, B. D. (2003). Stereotype threat and race of interviewer effects in a survey on political knowledge. *American Journal of Political Science*, *47*(1), 33–45. doi: 10.1111/1540-5907.00003

Dillman, D. A., & Christian, L. M. (2005). Survey mode as a source of instability in responses across surveys. *Field methods*, *17*(1), 30–52.

Fryer, Jr., R. G., & Levitt, S. J. (2004). The causes and consequences of distinctively Black names. *The Quarterly Journal of Economics*, *CXIX*(August), 767–806. doi: 10.1162/0033553041502180

Hatchett, S., & Schuman, H. (1975). White respondents and race-of-interviewer effects. *The Public Opinion Quarterly*, *39*(4), 523–528. doi: 10.1086/268249

Huddy, L., Billig, J., Bracciodieta, J., Moynihan, P. J., & Pugliani, P. (1997). The Effect of Interviewer Gender on the Survey Respnse. *Political Behavior*, *19*(3), 197–220. doi: 10.1023/A:1024882714254

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in cati, ivr, and web surveys the effects of mode and question sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865.

Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal Distance and Discrimination: An Audit Study in Academia. *Psychological Science*, *23*(7), 710–717. doi: 10.1177/0956797611434539

Reese, S. D., Danielson, W. A., Shoemaker, P. J., Chang, T.-K., & Hsu, H.-L. (1986). Ethnicity-of-Interviewer Effects Among Mexican-Americans and Anglos. *Public Opinion Quarterly*, *50*(4), 563–572. doi: 10.1086/269004

Rotnitzky, A., & Robins, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika*, *82*(4), 805–820.

Smith, R. A. (2002). Race, gender, and authority in the workplace: Theory and research. *Annual Review of Sociology*, 509–542.

Word, D. L., Coleman, C. D., Nunziata, R., & Kominski, R. (2008). Demographic aspects of surnames from census 2000. *Unpublished manuscript, Retrieved from http://citeseerx. ist. psu. edu/viewdoc/download*.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98. doi: 10.1007/s10683-009-9230-z