

Online Appendix for *Investigator Characteristics and Respondent Behavior in Online Surveys*

Contents

A Additional Reporting Standards Information 2

B Recruitment and Randomization Procedure 3

C Completion Rates 4

D Selection Into Treatment 6

E Attention Checks 9

F Power Calculations 10

G Exploratory Analysis of Effect Heterogeneity 11

H Survey Questions (Order Randomized) 13

I Pre-Analysis Plan 15

 I.1 Introduction: A Second Experiment on Researcher Identity in Online Surveys . . . 15

 I.2 Experimental Survey Design 16

 I.3 Statistical Tests 18

 I.4 Appendix: Survey Text 19

I.4.1	Informed Consent	19
I.4.2	Outcome Questions [Question order randomized]	21
J	A Second Experiment: Randomization in Only the Consent Form	27

A Additional Reporting Standards Information

Most of the requirements of the JEPS reporting standards are met by our main paper and various parts of this Appendix. In this section, we include several additional details that should aid in the understanding or replication of our study.

Recruitment Dates We recruited respondents and implemented the study between March 3 and April 4, 2016.

Incentives We paid survey participants between \$0.05 and \$1 for each wave of the study.

Evidence of Random Assignment Here we present group means of background covariates for respondents assigned to our different treatment conditions, as evidence of balance (suggesting our random assignment to treatment conditions worked as expected).

Variable	N	Male Researcher	Female Researcher	T-statistic
Prop. Male	2005	0.508	0.468	1.782
Prop. Democrat	2004	0.437	0.456	-0.839
Prop. Republican	2004	0.220	0.198	1.232
Prop. Voted in 2012	2005	0.703	0.729	-1.261
Prop. White	2002	0.820	0.805	0.837
Prop. Black	2002	0.070	0.065	0.389

Table A1: Covariate means by treatment conditions - Researcher gender

Variable	N	White Researcher	Black Researcher	T-statistic
Prop. Male	2005	0.484	0.491	-0.337
Prop. Democrat	2004	0.441	0.453	-0.544
Prop. Republican	2004	0.206	0.211	-0.307
Prop. Voted in 2012	2005	0.714	0.718	-0.191
Prop. White	2002	0.809	0.815	-0.365
Prop. Black	2002	0.074	0.061	1.170

Table A2: Covariate means by treatment conditions - Researcher race

Outcome Means/Standard Deviations Here we present means and standard deviations for all of our outcome measures.

Outcome	Mean	Std. Dev.	Number of Responses
Racial Resentment	0.468	0.280	1999
Would vote for a black president	0.934	0.249	2002
Would vote for a woman president	0.938	0.241	2003
Women’s role: Equal	0.904	0.295	2004
Women’s role: Home	0.053	0.225	2004
Expand public services	0.619	0.486	2005
Cut public services	0.381	0.486	2005

Table A3: Summary statistics for outcome variables

B Recruitment and Randomization Procedure

In order to assign Mechanical Turk workers to each of our treatment conditions, we first needed to obtain a sufficient number of candidate worker IDs that we could assign to each researcher name treatment. We recruited these respondents via an initial HIT which yielded a total of 5858 respondents. We then removed respondents that had either taken a previous version of one of our pilot studies, or took the survey multiple times. Doing so reduced the number of candidate respondents to 4659. We allocated respondents uniformly to the researcher name treatments in two waves, the first wave containing 3210 and the second wave containing 1449 respondents.

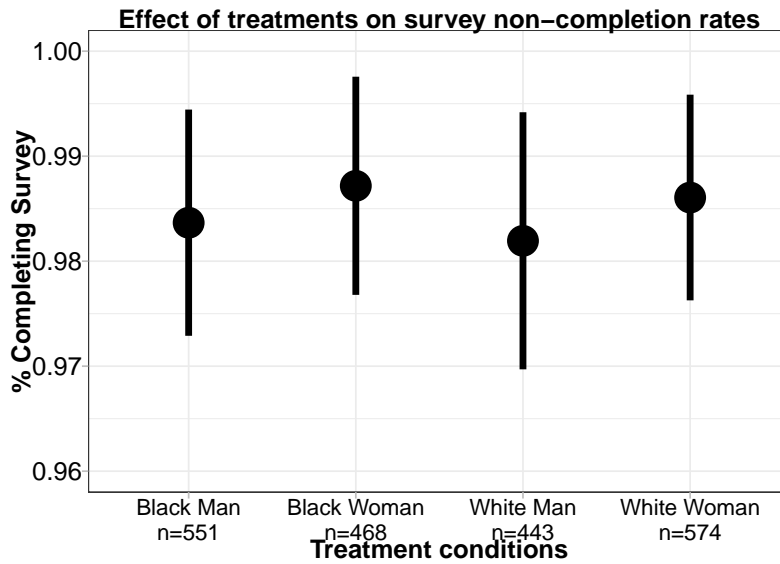
Of the 4659 respondents assigned to take our HIT, we received 2110 response to all versions of our survey. However, some of these respondents either accidentally or intentionally took the survey twice, due to a lag between the time that a new version of a HIT was posted and when we were able to remove a respondent from being qualified to take the HIT. We were able to recover MTurk IDs for most respondents and removed any duplicate HITs beyond the first taken by the respondent. For those responses where we could not record an MTurk ID, possibly due to respondents disabling some Javascript code execution on their browsers, we drop responses based on duplicated IP addresses. While this is a bit conservative as multiple users can in theory share the same IP address if they are on the same network (such as a public library), we want to avoid including multiple observations

from the same source. In all cases, the observations with duplicated IPs had the same treatment assignment, suggesting that we are picking up on the same person entering the survey multiple times. Our replication code provides the exact mechanism that we used to complete this de-duplication task. However, for privacy reasons, we do not release the IP addresses or the Mechanical Turk IDs of the respondents. We convert both IP address and MTurk ID to an arbitrary numerical indicator in the replication files that accompany this paper. After removing duplicates, we are left with 2036 unique observations. In total, 2016 of our 2036 respondents completed the survey and reached the end of the task. Of these 2016, we could match 2006 to one of the 4659 Mechanical Turk IDs we assigned at the beginning. We drop responses from one of these participants due to a request from that person after they completed the survey, for a total of 2005 responses. These 2005 constitute our sample of “completed” surveys.

C Completion Rates

An important element of non-response is the question of whether respondents are differentially failing to complete the survey based on which treatment condition they receive. We do not find strong evidence that this is the case. Figure [A1](#) plots the share of respondents under each treatment category that we could identify had completed the survey via Mechanical Turk. We fail to reject the null hypothesis of no differences in completion rates between each of the four treatment categories ($p > .05$). Also, in general, completion rates were very high (between 98 and 99 percent), so we do not have significant concerns over potential post-treatment bias induced by conditioning only on those respondents that finished the survey.¹

¹Note that in estimating these differences, we are conditioning on individuals selecting into the survey in the first place. We discuss selection into treatment in Section D below.



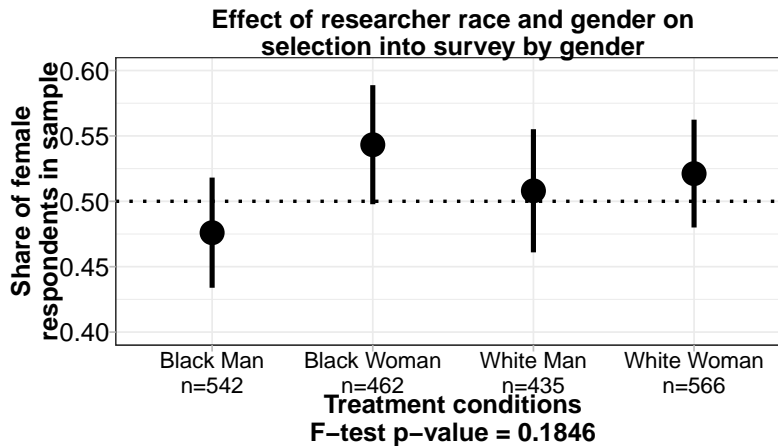
Note: Lines denote 95% confidence intervals.

Figure A1: Completion rates conditional on starting survey across four treatment categories.

D Selection Into Treatment

One of the mechanisms through which cues of researcher identity might affect survey results and responses is through differential selection into treatment. Some individuals may be more or less likely to take a survey after receiving information about the researcher, and differences in responses to survey questions may therefore be attributable to differences in respondents’ covariate distributions.

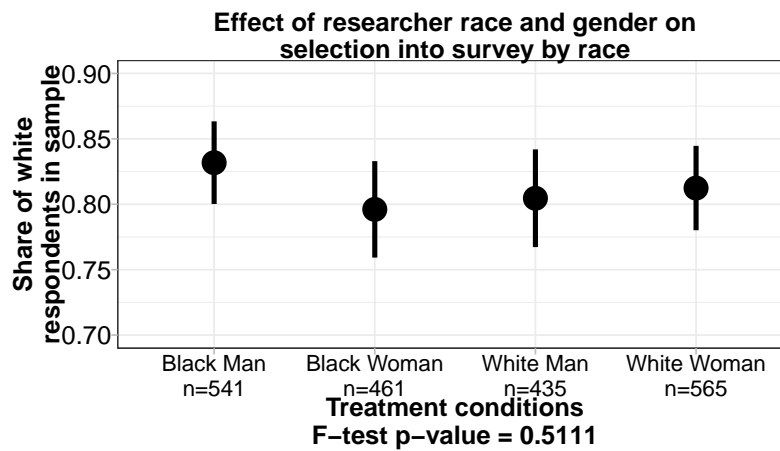
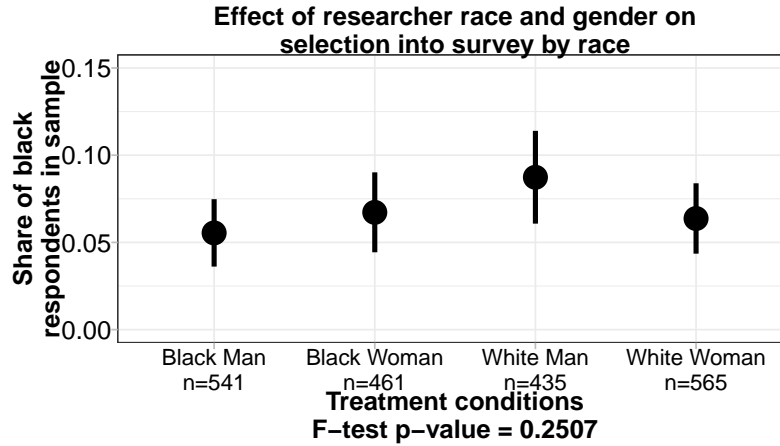
We do find evidence that some of our Mechanical Turk HITs were more likely to be taken by respondents. We test the null hypothesis that the observed counts in each of the four treatment categories are statistically indistinguishable from a uniform distribution by calculating a test statistic, the difference between the largest and smallest number of respondents in a treatment category. We then compared the observed test statistic to the null distribution where each respondent in the survey was assigned to each of the four treatment categories with probability .25. Using Monte Carlo simulation, we calculate a p-value of 0.00018, far below the common threshold of $\alpha = .05$. Based on the test, there is evidence to conclude that there may be some difference in the “popularity” of HITs based on the researcher name in the advertisement.



Note: Lines denote 95% confidence intervals.

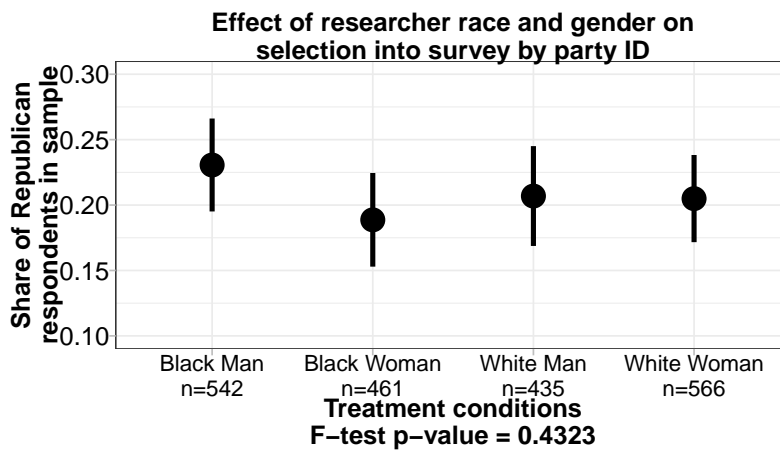
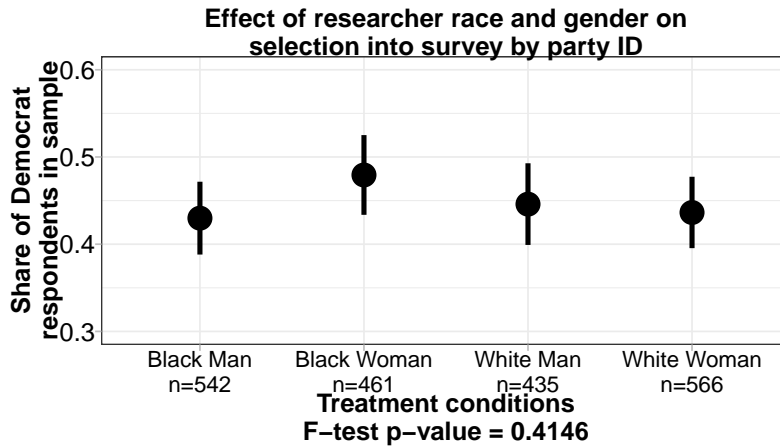
Figure A2: Differences in treatment groups across respondents’ background covariates: Gender

However, the magnitude of this difference is not particularly large. And more importantly, there do not appear to be significant differences between the samples for any of the four treatment categories based on the background covariates that we collected. Figure A2 plots the share of female



Note: Lines denote 95% confidence intervals.

Figure A3: Differences in treatment groups across respondents' background covariates: Race respondents in each of the four treatment categories. Using an F-test, we cannot reject the null of no difference across all four treatment categories (at $\alpha = .05$). The same null result holds for respondents who identify as white, for those who identify as black (Figure A3) and for political party identification (Figure A4). If there is some interesting selection into treatment, it is not clear that it relates to any of these commonly observed background covariates.

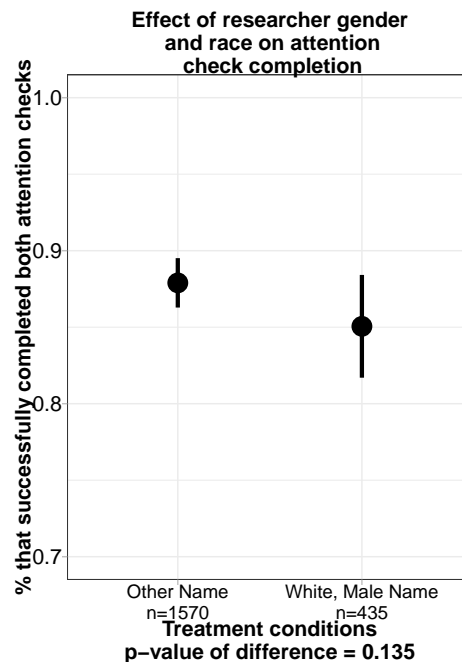


Note: Lines denote 95% confidence intervals.

Figure A4: Differences in treatment groups across respondents' background covariates: Party ID

E Attention Checks

Our final set of outcome questions test for whether respondents pay greater attention during surveys conducted by white male researchers when compared to all other treatment categories. We measure attention using two attention check questions randomly introduced in the experiment. Figure A5 reports the estimated probabilities of completing both attention checks correctly under the white male and non-white male treatments. Roughly 85 to 88 percent of respondents in both treatment groups successfully navigated the two attention check questions and the difference between the two is not statistically significant at $\alpha = .05$. Not only did treatment not appear to affect response substance (in terms of policy and attitude questions), it also did not significantly influence response quality (in terms of attention).



Note: Lines denote 95% confidence intervals.

Figure A5: Differences in attention check outcomes for researcher race/gender treatment.

F Power Calculations

This section outlines our procedure for evaluating the power of our hypothesis test of the effect of treatment on the three gender equality outcomes. To calculate the power of a hypothesis test, we need to make assumptions on the sampling distribution of the test statistics under the null. In our case, we have three large-sample difference-in-means tests between two proportions. While we fix the sample size ex-ante and vary the hypothetical “true” effect size for each treatment, we need to make an additional assumption about the population variance for each outcome. We start by using the sample variance of each of the three outcomes: beliefs about gender roles, willingness to vote for a woman president, and support for social service spending. Since we want to be somewhat conservative in our power analyses and account for potentially more variable samples, we scaled up each of the in-sample estimates by a factor of 1.5 (bounded by the maximum of .25 for a binary variable).

We evaluate power for two scenarios: one where all three nulls are false and have equal effect sizes, and one in which only one null is false (for the highest variance outcome). In the first scenario, we compute the “average power” (the share of nulls rejected) (Benjamini and Liu, 1999). In the second, we calculate the probability of rejecting the *non-null* hypothesis specifically. Figure A6 plots the two power curves for the hypothetical difference-in-means test between two proportions. We consider the case of $n = 2000$ evenly split between treatment and control. The y-axis denotes the power or expected power and the x-axis plots the hypothesized true absolute effect size. Following the Benjamini-Hochberg procedure for a false discovery rate of .05, we see that with a relatively high power of .8, we would detect effects greater than or equal to 5 percentage points. So even though we failed to reject the null hypothesis, the power of our design is such that it is unlikely that this would have happened had the true effects been of a substantial size (greater than 5%).

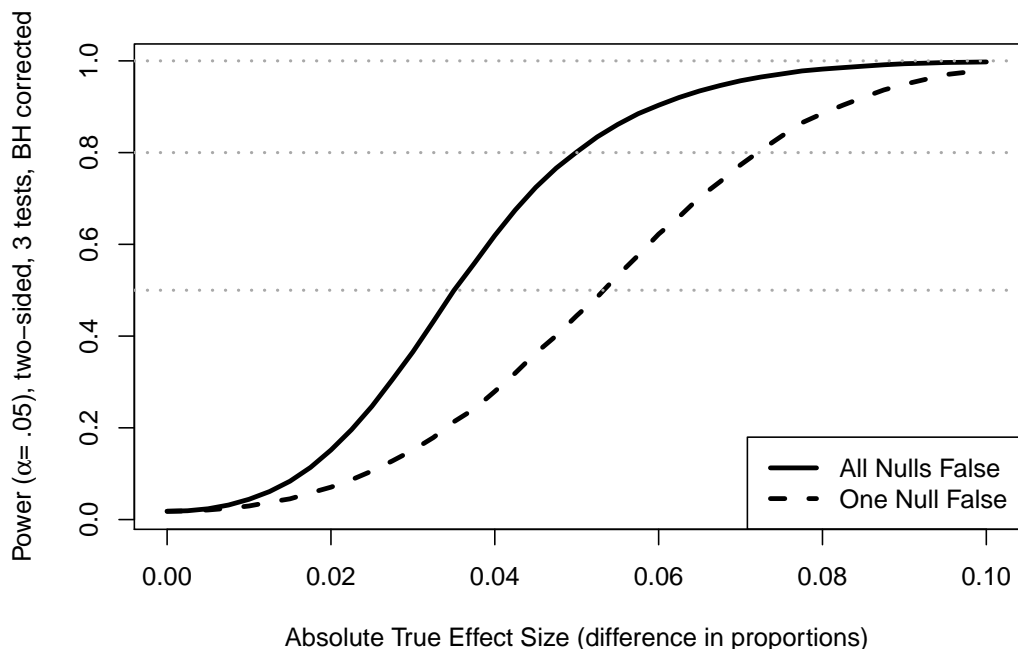
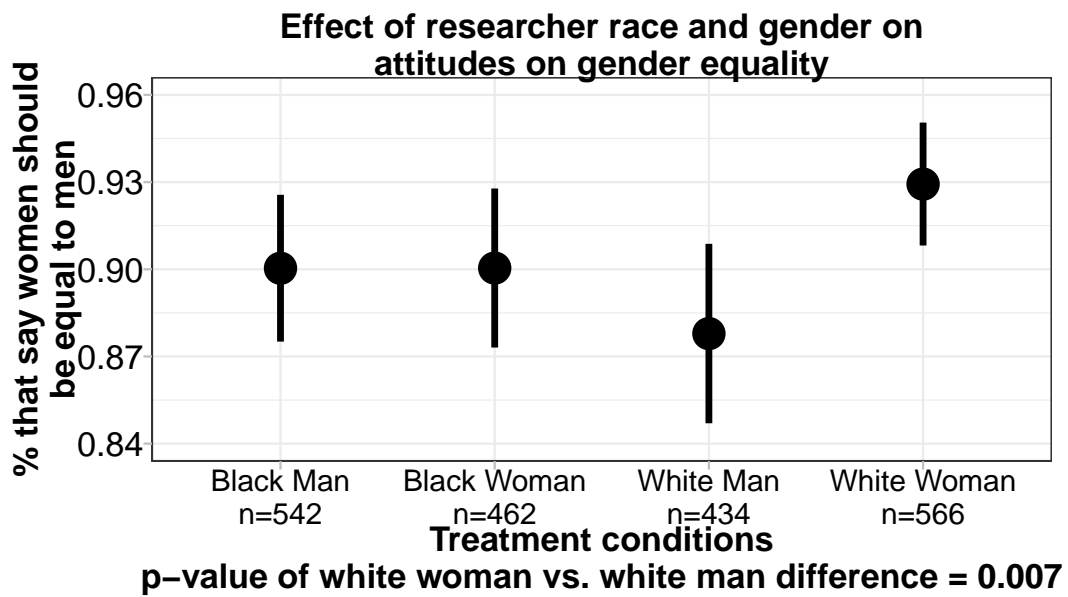


Figure A6: Power curves for three Benjamini-Hochberg adjusted large-sample two-sided difference in proportions tests. $n = 2000$ split evenly between groups.

G Exploratory Analysis of Effect Heterogeneity

Exploratory analyses provide some evidence of interesting potential effect heterogeneity when unpacking the female researcher effect on reported attitudes towards women’s equality. Specifically, we find interesting variation in race. It is important to note that the analysis that follows was not registered in our pre-analysis plan. While respondents in the putative white female researcher condition are significantly more likely to state that women should be equal to men in industry, business, and government relative to respondents in the white male researcher condition, we find no such effect of gender among the black researcher conditions. Figure A7 plots the estimated proportions of respondents in all four treatment categories. The p-value of the difference between the white woman and white man conditions is quite small at .007. However, since we did not pre-specify hypotheses about race and gender interactions when registering the study, this finding should be considered exploratory. However, based on its strength, we believe it warrants further investigation.



Note: Lines denote 95% confidence intervals.

Figure A7: Exploratory analyses for effect heterogeneity across race for gender effect on attitudes towards women's equality.

H Survey Questions (Order Randomized)

Closed-Ended Survey Questions

- In the past few years, we have heard a lot about improving the position of black people in this country. How much real change do you think there has been in the position of black people in the past few years: a lot, some, or not much at all?

Not much

Some

A lot

- 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

Women's place is in the home

Haven't 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

- While taking this survey, did you engage in any of the following behaviors? Please check all that apply.

Use your cell phone

Browse the internet

Talk with another person

Watch TV

Listen to music

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

1. Over the past few years, blacks have gotten less than they deserve. (agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly)

2. Irish, Italian, Jewish, and many other minorities overcame prejudice and worked their way up. Blacks should do the same without any special favors. (agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly)

3. 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. (agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly)
4. Generations of slavery and discrimination have created conditions that make it difficult for blacks to work their way out of the lower class. (agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly)

I Pre-Analysis Plan

I.1 Introduction: A Second Experiment on Researcher Identity in Online Surveys

The experiment discussed in this pre-analysis plan is the second component of a broader project exploring how the perceived identity of researchers affects responses in online surveys. In particular, we analyze how information about the race and gender of the researcher, conveyed in the required consent form and researcher username, affect survey responses. While the theoretical component and hypotheses across these two experiments are identical, this second experiment presents a slightly different way of testing the hypotheses by directly varying the researcher identity via the username used to post the survey.² This is in contrast to the first round of the experiment in which we used a generic username called “survey researcher” and only manipulated the researcher name on the consent form.

The reasons for explicitly manipulating the researcher username are two-fold. First, this is a common practice for researchers conducting online survey experiments. Indeed, several of the authors contributing to this paper explicitly list their name when advertising for their survey. It

²The first experiment, with a more complete discussion of the theory and hypotheses, was also pre-registered on the Experiments in Governance and Politics website here:

<http://egap.org/design-registration/registered-designs/>.

is useful to know whether information conveyed through this common practice can affect results, in order to establish best practices for how researchers should implement survey experiments on popular online survey platforms. Second, researcher usernames are displayed prominently on online survey platforms such as Amazon’s Mechanical Turk (MTurk). This means that survey respondents are more likely to be cognizant of the name of the researcher when proceeding from the survey posting to the survey itself. If, as we discuss in the related pre-registration document, we expect that perceived researcher identity can affect survey responses, then usernames represent one commonly used means through which respondents can learn the identity of the researcher.

I.2 Experimental Survey Design

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. In online recruitment platforms like Mechanical Turk, accounts have an associated username. We use this username to connote race and gender.

Because the treatment is in the username, rather than in the experiment, subject recruitment must be conducted carefully. In particular, the problem is that the treatment is no longer embedded in the survey (which makes randomization relatively simple) but instead in the request for users to take the survey. We solve this problem as follows.

We pool samples from previous studies conducted by the authors on Mechanical Turk. From this existing pool of respondents, we assign each to one of eight names, listed in Table A4. Subjects are then given a qualification associated with the condition to which they are assigned. Then, we create separate MTurk accounts for each username and deploy the same survey within each account. Each deployed survey is set such that only MTurk workers with the correct qualification can see the survey (and thus the username associated with it).

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 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 \(2004\)](#). We crossed the lists of first and last names to produce many possible combinations³ and drew two 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 A4.

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

Table A4: Names used for each of the four investigator name manipulations, based on lists from [Bertrand and Mullainathan \(2004\)](#), [Fryer, Jr. and Levitt \(2004\)](#), [Word et al. \(2008\)](#)

Within the survey, 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 H. 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, Margolis and Sances \(2014\)](#). 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.

³We omitted a few randomly-generated names that already belonged to celebrities, such as Jermaine Jackson.

I.3 Statistical Tests

For Hypothesis 1, we estimate two separate treatment effects.⁴ 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.

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 and Hochberg, 1995). This entails a two-step procedure where we order the 3 p-values of the individual hypothesis tests from smallest to largest, $p_{(1)}, \dots, p_{(3)}$ and then set our rejection level to $p_{(k)}$, where k is the largest value of

⁴Note that the hypotheses referred to in this section reference those laid out in the first preregistration document, which are as follows. H_1 : Assignment to a putatively female/black investigator will increase support for policies that provide for and protect the rights of women/blacks. H_2 : Attention and effort will be greatest among subjects assigned to a putatively white, male investigator.

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.

Additionally, because respondents have the option to stop taking the survey after treatment is assigned but before outcomes are measured, 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. Although it is not possible to adjust for nonignorable drop-out in the absence of prior covariates on respondents, we will examine whether there appears to be systematic differences between treatment arms with respect to attrition and employ sensitivity analyses in the vein of [Scharfstein, Rotnitzky and Robins \(1999\)](#) in order to evaluate the robustness of our estimates to this potential source of bias.

I.4 Appendix: Survey Text

I.4.1 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

jjjaeger@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 discon-

tinue participation at any time if I so choose. Answering yes continues the survey, answering no ends the survey.

Yes

No

I.4.2 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)

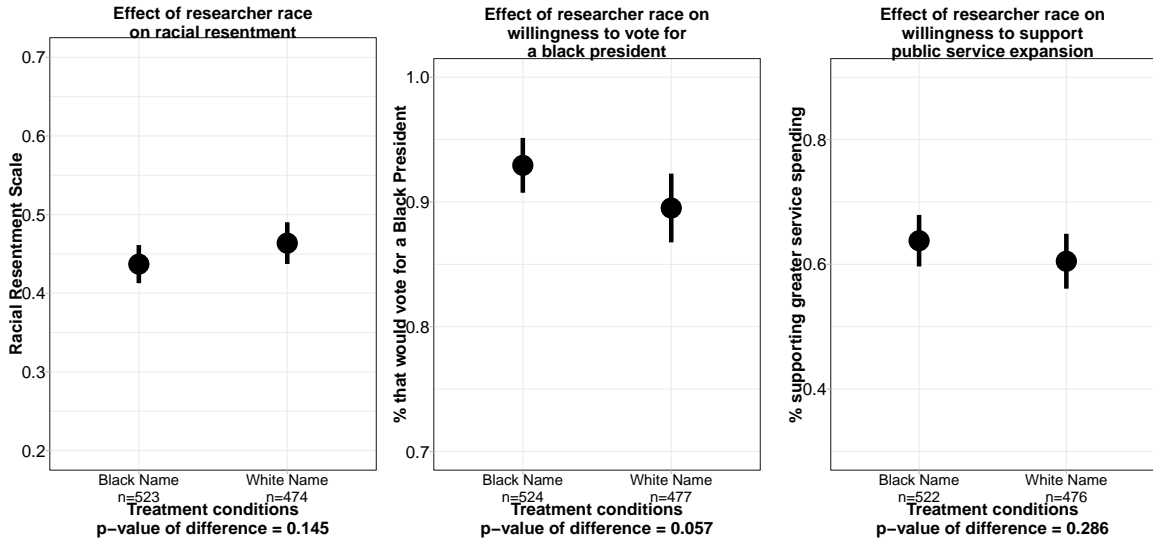
The purpose of this study was to learn about how researcher identity influences subjects' responses. In order to learn about this, we needed to manipulate the identity of the researcher. Though the contact information you were given was in fact real, the name of the researcher was not. This study was 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

J A Second Experiment: Randomization in Only the Consent Form

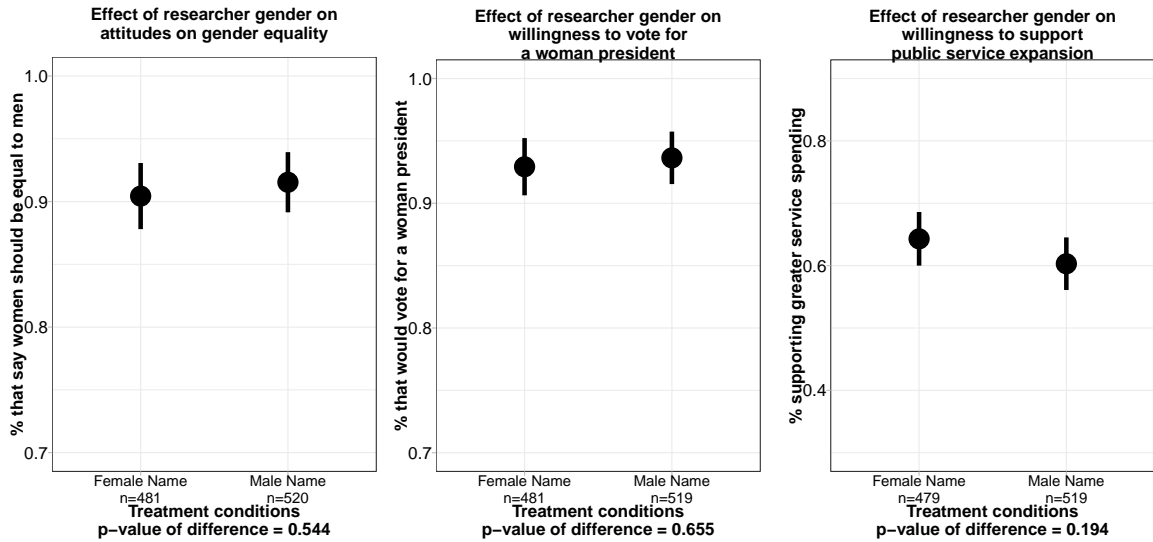
In addition to the experiment presented in this paper, we ran a second experiment in which we used a generic account name to post the initial HIT and then varied the researcher name only in the consent form.⁵ This allowed us to (1) test how respondents react to researcher identity when presented with a more subtle treatment, and (2) estimate the effect of researcher name on survey responses without the initial selection stage into the survey conditional on the name posted in the HIT. This second experiment was also fielded on Amazon's Mechanical Turk on a sample of 1000 respondents. We found no statistically discernible effect of treatment on any of the outcomes of interest. The full results for the pilot experiment are presented in Figures [A8-A11](#).

⁵This experiment was pre-registered on EGAP with study ID 20150717AA.



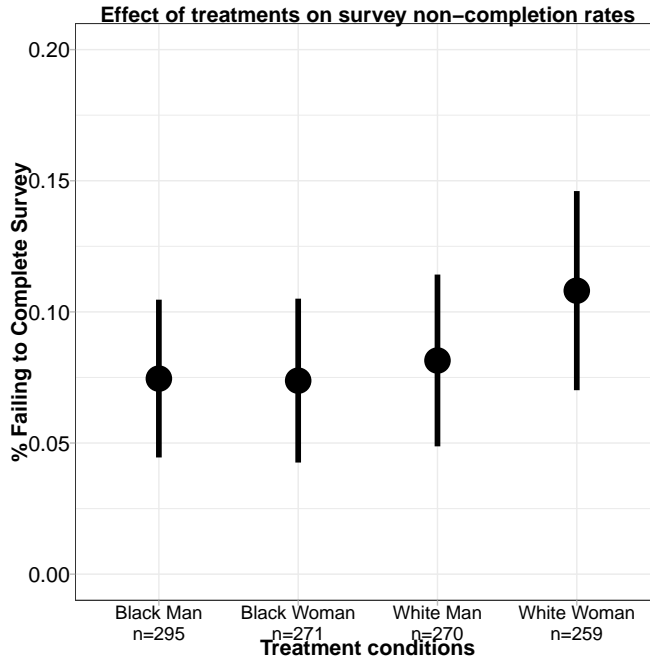
Note: Lines denote 95% confidence intervals.

Figure A8: Differences in policy/attitude outcomes for researcher race treatment in consent form only randomization experiment.



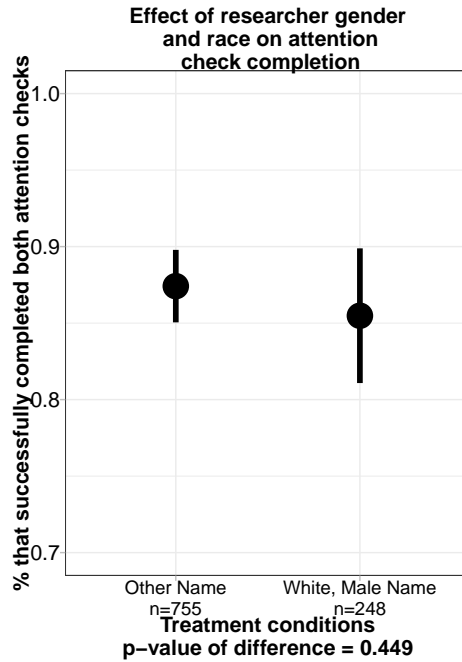
Note: Lines denote 95% confidence intervals.

Figure A9: Differences in policy/attitude outcomes for researcher gender treatment in consent form only randomization experiment.



Note: Lines denote 95% confidence intervals.

Figure A10: Estimated drop-out rates for experiment 1 across treatment conditions in consent form only randomization experiment.



Note: Lines denote 95% confidence intervals.

Figure A11: Differences in attention check outcomes for researcher race/gender treatment in consent form only randomization experiment.

References

- Benjamini, Yoav and Wei Liu. 1999. “A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence.” *Journal of Statistical Planning and Inference* 82(1):163–170.
- 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)* pp. 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.
- 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.
- Fryer, Jr., Roland G. and Steven J. Levitt. 2004. “The Causes and Consequences of Distinctively Black Names.” *The Quarterly Journal of Economics* CXIX(August):767–806.
- Scharfstein, Daniel O, Andrea Rotnitzky and James M Robins. 1999. “Adjusting for nonignorable drop-out using semiparametric nonresponse models.” *Journal of the American Statistical Association* 94(448):1096–1120.
- 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>.