Online Appendix for

"Misinformation and the Justification of Socially Undesirable Preferences"

D.J. Flynn*
Postdoctoral Fellow
Program in Quantitative Social Science
Dartmouth College
(603) 646-3995
d.j.flynn@dartmouth.edu

Yanna Krupnikov
Associate Professor
Department of Political Science
Stony Brook University

*Corresponding author

Appendix 1: Experimental Structure and CONSORT Participant Flow

This document was created following the experimental reporting standards outlined by JEPS (categories taken directly from the document available here: http://journals.cambridge.org/images/fileUpload/documents/xps_reportingstandards.pdf [accessed 8 February 2018]).

A. Hypotheses:
   a. Key Question: Under what conditions are participants more or less responsive to corrections of misinformation about minority candidates?
   b. Key Hypotheses:
      i. Individuals who are high in racial prejudice will be less responsive to a correction about a minority candidate of their own party than to a correction about a minority candidate of the opposing party.
      ii. Individuals who are high in racial prejudice will be more responsive to a correction about a white candidate of their own party than to a correction about a white candidate of the other party.

B. Subjects and Context:
   a. Participants were recruited online by Survey Sampling International (SSI). SSI maintains a panel of participants; from this panel, participants were recruited to participate in this particular study.
   b. Participants were eligible to participate if they were over 18 years old, were residents of the United States, and received a direct invitation from SSI (i.e., the invitation was not forwarded via email by an SSI panel member to a non-members of the SSI panel). Eligibility questions were asked at the very beginning of the study per SSI policy and IRB requirements. Individuals who did not meet eligibility criteria were immediately informed that they could not participate in the study.
   c. The study and recruitment all took place online. We used Qualtrics to record responses.
   d. The initial study took place between July 2-9, 2014. Based on input during the peer review process, we re-contacted our participants and invited them to take part in a follow-up study, which was fielded from February 1-8, 2016 (more details on re-contact study are provided below).

C. Allocation Method:
   a. Participants were randomized at the individual level using the Qualtrics randomization algorithm.
   b. Evidence of random assignment (both a randomization check and pre-treatment measures by condition) are included in Appendix 3A.

D. Treatments
   a. Full descriptions of treatments are provided in Appendix 2.
   b. The experiment was conducted entirely online via Qualtrics.
   c. Additional experimental details:
      i. This was a between-subject experiment

ii. No particular incentives were given for this study, though participants did earn SSI points as they would for any other study completed through SSI.

E. Results:

a. All measures are listed in Appendix 2.
b. Subgroup analyses:
    i. Analyses by racial attitudes were specified by hypotheses prior to the start of the experiment.
    ii. In responding to issues raised during the review process, additional subgroup analyses were conducted that were not initially specified by hypotheses: (1) an analysis by education levels was conducted as participants with higher levels of education may be more susceptible to social desirability pressures (e.g. Karp and Brockington 2005) and (2) analyses were conducted by party identification to ensure similar patterns across both parties.
    iii. In addition to the analyses reported in the manuscript, analyses with only white participants were also conducted; this was done as previous scholarship that uses the same prejudice measures largely focuses on white participants (Piston 2010; Tesler 2013). Results remain the same when black and Latino participants are excluded from the analysis.
c. Outcome analyses: All measures intended to be used as outcome measures are included in the presented analyses and in the appendix.
d. CONSORT participant flow diagram is presented in the next section of this appendix (Appendix 1). Given the number of conditions in the study, we present the detailed, by-group N in a separate table.
    i. Information about patterns of missing data is presented in Appendix 3B.
    ii. Information about the relationship between N and statistical significance is presented in Appendix 3C.
    iii. Covariate means by condition are discussed in Appendix 4, where we also elaborate on the racial prejudice measures.

F. Other Information

a. IRB approval obtained via Northwestern University. Approval for follow-up study obtained via Northwestern University.
b. All funding for the studies was provided by Northwestern University.
c. Replication dataset: https://dataverse.harvard.edu/dataverse.xhtml?alias=xps

CONSORT Participant Flow: Table by Treatment N

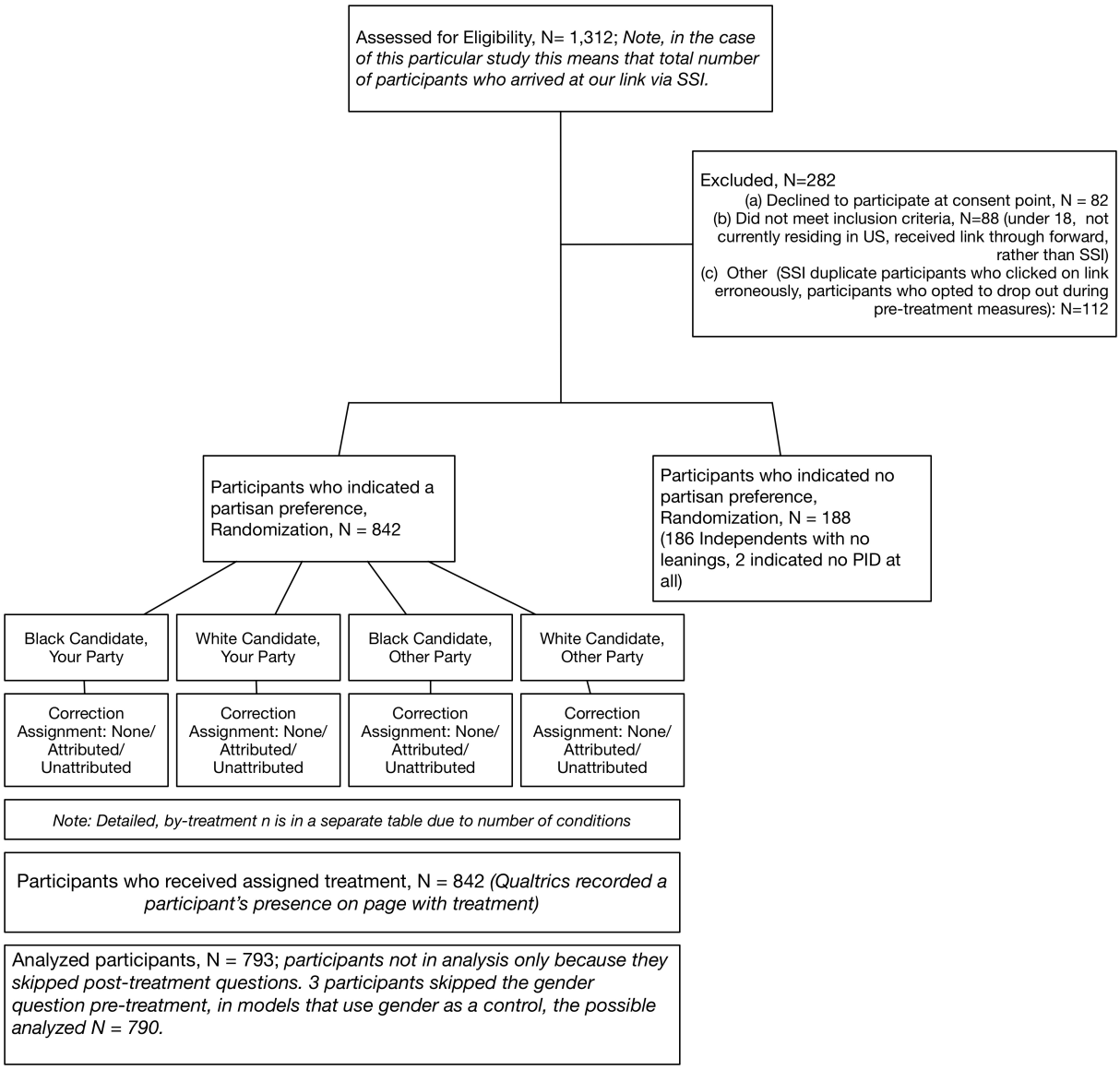

Table with Additional Details		
Enrollment	Assessed for Eligibility, N= 1,312	*Note, in the case of this particular study this means that total number of participants who arrived at our link via SSI.*
	Excluded, N = 282	Reasons for Exclusion: • Declined to participate at consent point, N = 82 • Did not meet inclusion criteria, N=88 (under 18, not currently residing in US, received link through forward, rather than SSI) • Other, N=112 (SSI duplicate participants who clicked on link erroneously for a second time, participants who opted to drop out during pre-treatment measures)
	Randomization, Participants who did not identify with a party/leaning toward a party, N=188	Given that our hypotheses were based on partisan attachments, we randomized people who did not identify with a party separately. We do so following previous research where partisan attachments are part of the hypothesis (see, e.g., Druckman et al. 2013). We randomize these individuals separately (rather than randomizing all participants at once and then excluding participants post-randomization from the analysis) following research that suggests exclusions should be made pre-randomization.
	Randomization, Participants identified with a party/leaning toward a party, N=842	
Given the differences in group N, we completed analyses to ensure that our results are not a function of group size. See Appendix 3, Table 3C. We also account for the differences using a model in Appendix 6.		
Allocation within partisans	Treatment 1, N = 60	Black candidate, P's party, no correction
	Treatment 2, N = 52	Black candidate, P's party, attributed correction
	Treatment 3, N = 50	Black candidate, P's party, unattributed correction
	Treatment 4, N = 100	White candidate, P's party, no correction
	Treatment 5, N = 59	White candidate, P's party, attributed correction
	Treatment 6, N = 80	White candidate, P's party, unattributed correction
	Treatment 7, N = 61	Black candidate, other party, no correction
	Treatment 8, N = 48	Black candidate, other party, attributed correction
	Treatment 9, N = 73	Black candidate, other party, unattributed correction
	Treatment 10, N = 92	White candidate, other party, no correction
	Treatment 11, N = 60	White candidate, other party, attributed correction
	Treatment 12, N = 107	White candidate, other party, unattributed correction
Analysis within	Treatment 1, N = 57	Black candidate, P's party, no correction
	Treatment 2, N = 52	Black candidate, P's party, attributed correction

partisans	Treatment 3, N = 48	Black candidate, P's party, unattributed correction
	Treatment 4, N = 86	White candidate, P's party, no correction
	Treatment 5, N = 58	White candidate, P's party, attributed correction
	Treatment 6, N = 78	White candidate, P's party, unattributed correction
	Treatment 7, N = 56	Black candidate, other party, no correction
	Treatment 8, N = 46	Black candidate, other party, attributed correction
	Treatment 9, N = 67	Black candidate, other party, unattributed correction
	Treatment 10, N = 86	White candidate, other party, no correction
	Treatment 11, N = 56	White candidate, other party, attributed correction
	Treatment 12, N = 100	White candidate, other party, unattributed correction
	All exclusions are due to participants not answering post-treatment questions	

Appendix 2: Full Text of Experimental Treatments and Measures

[All participants]

The State Capitol – This week the legislature continues its consideration of the education reform bill, with the final vote expected sometime in the next two weeks.

However, for one legislator, debate on the education package continues to be overshadowed by allegations that he used his position on a key house committee to secure a tax loophole for a large retailer headquartered in his district.

According to reports, Rep. Sam Larson **[(D)/(R)]** was instrumental in passing the amendment, which would significantly reduce the retailer's tax liability next year. Campaign finance records indicate that the CEO of the store contributed the maximum allowable amount to Larson's campaign last year. Larson's office did not respond to requests for comment.

[Attributed correction:]
According to PolitiCheck, a non-partisan fact-checking organization whose goal is to adjudicate factual disputes, Rep. Sam Larson **[(D)/(R)]** did not use his position on a key House committee to secure a tax loophole for a large retailer headquartered in his district. State records indicate that a representative from another part of the state introduced the measure in committee and secured its passage. Also, the company issued a press release stating that the CEO has never donated to Larson's campaign. The CEO has a common surname (Thomas), and a review of campaign finance records confirms that a different individual of the same name made the contributions.

[Unattributed correction:]
New evidence contradicts earlier reports that Rep. Sam Larson **[(D)/(R)]** used his position on a key House committee to secure a tax loophole for a large retailer headquartered in his district. State records indicate that a representative from another part of the state introduced the measure in committee and secured its passage. Also, the company issued a press release stating that the CEO has never donated to Larson's campaign. The CEO has a common surname (Thomas), and a review of campaign finance records confirms that a different individual of the same name made the contributions.

Candidate Pictures (Race Manipulation)

Measures:

Dependent Variables:

How favorable or unfavorable do you feel toward Rep. Larson?
-Very Favorable [7]
-Somewhat Favorable [6]
-Slightly Favorable [5]
-Neither Favorable Nor Unfavorable [4]
-Slightly Unfavorable [3]
-Somewhat Unfavorable [2]
-Very Unfavorable [1]

If Rep. Larson were up for re-election, how likely would you be to vote for him?
-I would be highly likely to vote for him [5]
-I would be somewhat likely to vote for him [4]
-I am uncertain if I would vote for him [3]
-I would be somewhat unlikely to vote for him [2]
-I would be highly unlikely to vote for him [1]

How good of a job do you think that Rep. Larson would probably do with the following…Be an effective Governor?

1 (do a very poor job) ---------- 7 (do an excellent job)

How good of a job do you think that Rep. Larson would probably do with the following…Be an effective U.S. President about 10 years from now?

1 (do a very poor job) ---------- 7 (do an excellent job)

How good of a job do you think that Rep. Larson would probably do with the following…Be an effective leader for his constituents?

1 (do a very poor job) ---------- 7 (do an excellent job)

Racial Attitudes Battery:

We would like you to place each group on a 7-point scale based on how well you think a particular characteristic applies to people in that group.

In the first statement, a score of '1' means that you think almost all of the people in that group tend to be "lazy." A score of '7' means that you think most people in the group are "hardworking." A score of '4' means that you think that most people in the group are not closer to one end or the other, and of course, you may choose any number in between.

Whites: 1 (lazy) ---------- 7 (hardworking)

Blacks: 1 (lazy) ---------- 7 (hardworking)

Whites: 1 (unintelligent) ---------- 7 (intelligent)

Blacks: 1 (unintelligent) ---------- 7 (intelligent)

Table 2A1: Sample Demographics

Variable	
% Democrat	37.5
% Republican	24.9
% with BA or higher	41.5
% Male	44.1
Median Income	$40,000-$50,000

Appendix 3: Randomization, Missing Data, Cell Sizes, and Covariate Means

Appendix 3A: Randomization

We conducted a number of randomization checks, including model predicting group assignment by covariates and balance tests. Below we present the results of all the various tests. In the randomization checks that rely on models we focus on the assignment by candidate race and party within either the no correction or correction conditions. When we rely on other types of balance checks we focus on the candidate race assignment first.

3A.1. Models of Randomization

We first rely on a series of multinomial logits to consider likelihood of assignment to a certain group. We estimate these models with the racial stereotype variables included individually and as a full combined scale. We see no evidence that any of the variables predict particular assignment (see Table 3A1). Since we pool the correction source conditions, we do so in the models below as well.

Table 3A1: Multinomial Logit	
	Chi-square (p-value)
Partisanship	4.99 (p=0.6616)
White (1 if white, 0 otherwise)	3.17 (p=0.8692)
Income (1 low income to 9 high income)	5.91 (p=0.5509)
Gender (1-0)	6.40 (p=0.4933)
Racial Prejudice: Lazy	9.55 (p=0.2154)
Racial Prejudice: Intelligent Scale	9.09 (p=0.2461)
Racial Prejudice: Full Scale	7.98 (p=0.1323)

3A.2 Balance Tests

We also conduct a series of Kolmogorov-Smirnov tests to track balance across conditions. In particular, we focus on our measures of prejudice. These tests do not suggest that there are any differences across groups on measures of racial prejudice

(Table 3A2). Nonetheless, the test in 3A2 is done by condition; a secondary group comparison test (Bonferroni) leads to similar findings.

In addition to tests that focus on prejudice we also conducted a similar balance test on race as this factor too can influence outcomes. Again, we see no balance differences across conditions (Table 3A3).

Table 3A2: Racial Prejudice Scale							
	2	3	4	5	6	7	8
1	D= 0.04 p=1.000	D=0.08 p=0.978	D=0.07 p=0.975	D=0.15 p=0.499	D=0.06 p=1.000	D=0.05 p=1.000	D=0.10 p=0.726
2		D=0.09 p=0.778	D=0.05 p=0.999	D=0.14 p=0.423	D=0.05 p=0.999	D=0.06 p=0.990	D=0.12 p=0.326
3			D=0.05 p=1.000	D=0.11 p=0.784	D=0.08 p=0.886	D=0.07 p=0.984	D=0.06 p=0.965
4				D=0.11 p=0.680	D=0.08 p=0.830	D=0.05 p=1.00	D=0.08 p=0.774
5					D=0.15 p=0.356	D=0.12 p=0.689	D=0.11 p=0.672
6						D=0.06 p=0.982	D=0.09 p=0.650
7							D=0.09 p=0.714

Table 3A3: Race							
	2	**3**	**4**	5	6	**7**	**8**
1	D =0.05 p=1.000	D =0.06 p=1.000	D =0.04 p=1.000	D=0.005 p=1.000	D =0.02 p=1.000	D =0.04 p=1.000	D =0.003 p=1.000
2		D=0.02 p=1.000	D=0.003 p=1.000	D=0.06 p=0.999	D=0.05 p=0.999	D=0.04 p=1.000	D=0.02 p=1.000
3			D=0.03 p=1.000	D=0.06 p=0.999	D=0.05 p=0.999	D=0.04 p=1.000	D=0.03 p=1.000
4				D=0.03 p=1.000	D=0.04 p=1.000	D=0.05 p=1.000	D=0.01 p=1.000
5					D=0.05 p=1.000	D=0.08 p=0.983	D=0.05 p=1.000
6						D=0.03 p=1.000	D=0.03 p=1.000
7							D=0.03 p=1.000

Tables 3A2 and 3A3 rely on a Kolmogorov-Smirnov test. The null hypothesis in these comparisons is that there is no difference between groups. The p-values in each of these cases shows that the null cannot be rejected.

Appendix 3B: Tests of Missing Data

To consider whether there are any patterns to missing data we conduct several analyses. First, we track patterns in missing variables in each of the key variables we use: prejudice measures and outcome variables. Second, we create a joint measure for missing data (i.e. if a participant skipped any questions at all). We present all results in Table 3B1, which demonstrates no patterns to missing values in post-treatment data.

Table 3B1: Predicting Missing Values in Post-Treatment Data						
	Prejudice Scale (Missing 1, else 0)	Favorability (Missing 1, else 0)	Governing (Missing=1, else 0)	President (Missing 1, else 0)	Overall (Any Missing 1, else 0)	Count of missing responses
PID	0.008 (0.055)	-0.044 (0.037)	0.036 (0.037)	0.016 (0.033)	0.009 (0.030)	0.006 (0.043)
Gender	0.161 (0.265)	-0.085 (0.167)	0.006 (0.176)	-0.101 (0.155)	-0.054 (0.141)	-0.088 (0.202)
Candidate Treatment	-0.122 (0.317)	0.171 (0.229)	0.205 (0.241)	0.115 (0.218)	0.191 (0.195)	0.260 (0.269)
Correction	-0.308 (0.456)	0.289 (0.229)	0.225 (0.318)	0.151 (0.285)	0.222 (0.255)	0.330 (0.347)
Cand x Correction	0.084 (0.150)	-0.101 (0.101)	-0.071 (0.106)	-0.020 (0.094)	-0.055 (0.085)	-0.090 (0.117)
Constant	-1.889 (0.938)	-2.119 (0.660)	-2.659 (0.737)	-2.247 (0.660)	-2.239 (0.593)	2.991 (0.813)
χ^2 *(p-value)*	*1.10 (0.9543)*	*3.08 (0.6877)*	*2.07 (0.8393)*	*2.95 (0.7082)*	*2.61 (0.7597)*	*1.86 (0.8680)*

Models where missing is coded as 1 and all else as 0 are estimated using probit. The final model which uses the full count of missings as a dependent variable is estimated using Poisson. Note, that constitutive terms of the interaction term are difficult to interpret directly as they offer the effect of one term when the other is at 0. Since neither constitutive term takes on 0, direct interpretation is difficult.

Appendix 3C: Randomization and Size of Groups

As our CONSORT diagram demonstrates, we do have some uneven random assignment by group. Therefore, we track whether the differences in N could we responsible for the observed patterns. In particular, we focus on the candidate of the participant's own party. One pattern that we observe is that we see null effects for the Black candidate of the participant's own party, but significant effects for the White candidate of the participant's own party. To ensure this is not a function of cell size, we consider whether the size and variance of the effect observed for the Black candidate would have reached significance using the N obtained for the White candidate. Similarly, we track whether the effect observed for the White candidate would have reached significance given the N obtained for the Black candidate. These results are for high prejudice participants.

Since we rely on t-tests in our main analyses, we use t-tests here as well.

Table 3C1: Comparisons of cell size and effect		
Same Party, Black Candidate		
	Original Result	Same effect, N for Same Party, **White** Candidate
Favorability	0.18 (p=0.67)	0.18 (p=0.57)
Vote Willingness	-0.17 (p=0.59)	-0.17 (p=0.48)
Gubernatorial Success	0.20 (p=0.64)	0.20 (p=0.54)
Presidential Success	-0.04 (p=0.92)	-0.04 (p=0.88)
Leadership	0.21 (p=0.64)	0.22 (p=0.49)
Same Party, White Candidate		
	Original Result	Same effect, N for Same Party, **Black** Candidate
Favorability	1.35(p<0.001)	1.35 (0.002)
Vote Willingness	0.61 (p=0.02)	0.61 (p=0.09)
Gubernatorial Success	1.30 (p<0.001)	1.30 (p=0.002)
Presidential Success	1.36 (p<0.001)	1.36 (p=0.003)
Leadership	1.38 (p=0.0001)	1.38 (p=0.002)

Appendix 4: Condition Means and Comparisons

In the tables below we present condition means, difference in difference analyses and comparisons by race of the candidate. Difference-in-difference analyses are conducted following code provided by Jerit, Barabas, and Clifford (2013). Analyses provided for all outcome measures. Details on measures are provided in Appendix 2.

In appendix 4A we present the basic means and standard deviations, in appendix 4B we present results with p-values and difference-in-difference patterns, in appendix 4C we present additional comparisons for participant with negative racial attitudes.

Appendix 4A: Means and standard deviations by measure

Table 4A1: Mean Favorability Ratings by Experimental Condition, with Breakdowns by Prejudice Level

	All Participants		Low Prejudice Participants		High Prejudice Participants	
	No Correction	Correction	No Correction	Correction	No Correction	Correction
Same Party, Black	3.52 (1.42)	4.11 (1.29)	3.38 (1.35)	4.16 (1.33)	3.78 (1.59)	3.96 (1.19)
Same Party, White	3.18 (1.62)	4.24 (1.45)	3.32 (1.79)	4.16 (1.44)	3.03 (1.27)	4.38 (1.47)
Other Party, Black	2.57 (1.23)	3.71 (1.43)	2.82 (1.33)	3.91 (1.44)	2.22 (1.00)	3.29 (1.33)
Other Party, White	2.91 (1.51)	3.15 (1.60)	3.21 (1.55)	3.20 (1.64)	2.33 (1.27)	3.11 (1.56)

Note: Cell entries are mean favorability ratings (1-7 scale), with standard deviations in parentheses. Higher values mean higher favorability.

Table 4A2: Mean Likelihood of Voting for Candidate by Experimental Condition, with Breakdowns by Prejudice Level

	All Participants		Low Prejudice Participants)		High Prejudice Participants	
	No Correction	Correction	No Correction	Correction	No Correction	Correction
Same Party, Black	2.80 (1.02)	3.01 (1.04)	2.71 (0.98)	3.08 (1.01)	3.00 (1.14)	2.83 (1.10)
Same Party, White	2.56 (1.26)	3.08 (1.14)	2.56 (1.31)	3.04 (1.12)	2.54 (1.12)	3.14 (1.21)
Other Party, Black	1.97 (1.20)	2.50 (1.16)	2.40 (1.31)	2.61 (1.20)	1.42 (0.72)	2.26 (1.03)
Other Party, White	2.22 (1.27)	2.31 (1.26)	2.40 (1.32)	2.36 (1.33)	1.88 (1.10)	2.23 (1.15)

Note: Cell entries are mean values of reported likelihood of voting for Larson (1-5 scale), with standard deviations in parentheses. Higher values mean higher likelihood of voting.

Table 4A3: Mean Effective Governor Ratings by Experimental Condition, with Breakdowns by Prejudice Level

	All Participants		Low Prejudice Participants		High Prejudice Participants	
	No Correction	Correction	No Correction	Correction	No Correction	Correction
Same Party, Black	3.70 (1.34)	4.19 (1.23)	3.70 (1.32)	4.30 (1.16)	3.67 (1.46)	3.89 (1.37)
Same Party, White	3.40 (1.66)	4.25 (1.40)	3.60 (1.84)	4.14 (1.36)	3.17 (1.28)	4.47 (1.46)
Other Party, Black	2.79 (1.38)	3.72 (1.40)	3.03 (1.49)	3.94 (1.46)	3.42 (1.16)	3.24 (1.14)
Other Party, White	3.09 (1.54)	3.33 (1.57)	3.39 (1.56)	3.36 (1.63)	2.50 (1.33)	3.31 (1.50)

Note: Cell entries are mean "effective governor" ratings (1-7 scale), with standard deviations in parentheses. Higher values mean higher perceived effectiveness.

Table 4A4: Mean Effective President Ratings by Experimental Condition, with Breakdowns by Prejudice Level

	All Participants		Low Prejudice Participants		High Prejudice Participants	
	No Correction	Correction	No Correction	Correction	No Correction	Correction
Same Party, Black	3.31 (1.49)	3.72 (1.35)	3.33 (1.47)	3.93 (1.34)	3.22 (1.59)	3.18 (1.25)
Same Party, White	3.08 (1.69)	3.98 (1.54)	3.31 (1.83)	3.90 (1.54)	2.77 (1.37)	4.13 (1.53)
Other Party, Black	2.39 (1.30)	3.25 (1.43)	2.76 (1.41)	3.52 (1.48)	1.87 (0.92)	2.67 (1.12)
Other Party, White	2.89 (1.67)	3.01 (1.57)	3.22 (1.63)	3.10 (1.61)	2.26 (1.57)	2.91 (1.53)

Note: Cell entries are mean "effective president" ratings (1-7 scale), with standard deviations in parentheses. Higher values mean higher perceived effectiveness.

Table 4A5: Mean Leadership Ratings by Experimental Condition, with Breakdowns by Prejudice Level

	All Participants		Low Prejudice Participants		High Prejudice Participants	
	No Correction	Correction	No Correction	Correction	No Correction	Correction
Same Party, Black	3.73 (1.38)	4.09 (1.39)	3.80 (1.36)	4.21 (1.32)	3.57 (1.43)	3.79 (1.52)
Same Party, White	3.35 (1.71)	4.26 (1.50)	3.51 (1.94)	4.13 (1.45)	3.13 (1.32)	4.51 (1.58)
Other Party, Black	2.68 (1.43)	3.81 (1.52)	3.03 (1.57)	4.04 (1.63)	2.17 (1.03)	3.34 (1.15)
Other Party, White	3.16 (1.69)	3.37 (1.54)	3.42 (1.71)	3.40 (1.48)	2.62 (1.52)	3.32 (1.63)

Note: Cell entries are mean perceived leadership ratings (1-7 scale), with standard deviations in parentheses. Higher values mean higher perceived effectiveness.

Appendix 4B: Difference in Difference Comparisons

In the next set of tables we conduct t-test comparisons and difference-in-difference (D-I-D) comparisons across the conditions. D-I-D comparisons are done using code provided by Jerit, Barabas, and Clifford (2013). In the analyses below all p-values are based on two-tailed tests.

Table 4B1: Mean Favorability Ratings by Experimental Condition, with Breakdowns by Prejudice Level and D-I-D Comparisons

	Low Prejudice Participants			High Prejudice Participants		
	No Correction	Correction	Difference	No Correction	Correction	Difference
Same Party, Black	3.38 (1.35)	4.16 (1.33)	0.78 (p=0.004)	3.78 (1.59)	3.96 (1.19)	0.17 (p=0.6698)
Same Party, White	3.32 (1.79)	4.16 (1.44)	0.84 (p=0.002)	3.03 (1.27)	4.38 (1.47)	1.35 (p<0.001)
		D-I-D	0.06 (p=0.8664)		D-I-D	1.18 (p=0.0159)
Other Party, Black	2.82 (1.33)	3.91 (1.44)	1.09 (p=0.003)	2.22 (1.00)	3.29 (1.33)	1.07 (p=0.002)
Other Party, White	3.21 (1.55)	3.20 (1.64)	-0.01 (p=0.9674)	2.33 (1.27)	3.10 (1.56)	0.77 (p=0.02)
		D-I-D	1.10 (p=0.0055)		D-I-D	-0.30 (p=0.4953)

Table 4B2: Mean Likelihood of Voting for Candidate by Experimental Condition, with Breakdowns by Prejudice Level and D-I-D comparisons

	Low Prejudice Participants			High Prejudice Participants		
	No Correction	Correction	Difference	No Correction	Correction	Difference
Same Party, Black	2.71 (0.98)	3.08 (1.01)	0.37 (p=0.0573)	3.00 (1.14)	2.83 (1.10)	-0.17 (p=0.5995)
Same Party, White	2.56 (1.31)	3.04 (1.12)	0.49 (p=0.0152)	2.54 (1.12)	3.14 (1.21)	0.58 (p=0.0262)
		D-I-D	0.11 (p=0.6962)		D-I-D	0.76 (p=0.0758)
Other Party, Black	2.40 (1.31)	2.61 (1.20)	0.21 (p=0.3895)	1.42 (0.72)	2.26 (1.03)	0.88 (p=0.0003)
Other Party, White	2.40 (1.32)	2.36 (1.33)	-0.04 (p=0.8633)	1.88 (1.10)	2.23 (1.15)	0.36 (p=0.1477)
		D-I-D	-0.25 (p=0.4464)		D-I-D	-0.52 (p=0.1422)

Table 4B3: Mean Effective Governor Ratings by Experimental Condition, with Breakdowns by Prejudice Level and D-I-D Comparisons

	Low Prejudice			High Prejudice		
	No Correction	Correction	Difference	No Correction	Correction	Difference
Same Party, Black	3.70 (1.32)	4.30 (1.16)	0.60 (p=0.0138)	3.68 (1.46)	3.89 (1.37)	0.21 (p=0.6157)
Same Party, White	3.60 (1.84)	4.14 (1.36)	0.54 (p=0.0403)	3.11 (1.28)	4.47 (1.46)	1.35 (p<0.001)
		DID	-0.06 (p=0.8741)		DID	1.15 (p=0.027)
Other Party, Black	3.03 (1.49)	3.94 (1.46)	0.91 (p=0.0034)	2.43 (1.16)	3.24 (1.14)	0.81 (p=0.0103)
Other Party, White	3.39 (1.56)	3.36 (1.63)	-0.03 (p=0.8994)	2.50 (1.33)	3.27 (1.50)	0.77 (p=0.0174)
		DID	-0.94 (p=0.0205)		DID	-0.03 (p=0.9415)

Table 4B4: Mean Effective President Ratings by Experimental Condition, with Breakdowns by Prejudice Level and D-I-D Comparisons

	Low Prejudice Participants			High Prejudice Participants		
	No Correction	Correction	Difference	No Correction	Correction	Difference
Same Party, Black	3.33 (1.47)	3.93 (1.34)	0.60 (p=0.0317)	3.26 (1.59)	3.18 (1.25)	-0.08 (p=0.8376)
Same Party, White	3.31 (1.83)	3.90 (1.54)	0.59 (p=0.374)	2.77 (1.37)	4.13 (1.53)	1.36 (p<0.0001)
		DID	-0.01 (p=0.9897)		DID	1.44 (p=0.0054)
Other Party, Black	2.76 (1.41)	3.52 (1.48)	0.76 (p=0.0136)	1.87 (0.92)	2.67 (1.12)	0.80 (p=0.006)
Other Party, White	3.22 (1.63)	3.10 (1.61)	-0.12 (p=0.6457)	2.26 (1.57)	2.86 (1.53)	0.61 (p=0.0759)
		DID	0.88 (p=0.0307)		DID	-0.19 (p=0.6936)

Table 4B5: Mean Leadership Ratings by Experimental Condition, with Breakdowns by Prejudice Level and D-I-D Comparisons

	Low Prejudice Participants			High Prejudice Participants		
	No Correction	Correction	Difference	No Correction	Correction	Difference
Same Party, Black	3.8 (1.36)	4.21 (1.32)	0.41 (p=0.1259)	3.57 (1.43)	3.79 (1.52)	0.21 (p=0.6419)
Same Party, White	3.51 (1.94)	4.13 (1.45)	0.62 (p=0.030)	3.13 (1.32)	4.51 (1.58)	1.38 (p=0.0001)
		DID	0.42 (p=0.5775)		DID	1.17 (p=0.0356)
Other Party, Black	3.03 (1.57)	4.04 (1.63)	1.01 (p=0.003)	2.17 (1.03)	3.34 (1.15)	1.16 (p=0.0002)
Other Party, White	3.42 (1.71)	3.40 (1.48)	-0.02 (p=0.9269)	2.62 (1.52)	3.32 (1.63)	0.70 (p=0.0536)
		DID	-1.008 (p=0.0139)		DID	-0.47 (p=0.3518)

Appendix 4C: Comparisons within correction condition

We conduct a final set of comparisons within the correction condition by race of candidate. Our goal here is to consider whether participants with negative racial attitudes are willing to go to higher ends of the scale for white rather than black candidates.

We observe the following patterns in Table 4C1: participants who have more negative racial attitudes are willing to go to higher ends of the scale for the black candidate than for the white candidate. We can observe this effect in two ways. First, the higher rating for the white candidate of the same party than the black candidate of the same party in the corrected condition. Second, the significant D-I-D value for the same party candidates between the uncorrected and the corrected conditions. If shifts were simply proportional to where the candidates started in the uncorrected conditions, the D-I-D would not reflect differences.

Table 4C1			
	Uncorrected, Δ Race	Corrected, Δ Race	D-I-D
Favorability			
Same Party	-0.82 (p=0.04)	0.41 (p=0.21)	1.22 (p=0.0180)
Other Party	-0.12 (p=0.7198)	0.21 (p=0.4846)	0.32 (p=0.4947)
Vote			
Same Party	-0.44 (p=0.18)	0.32 (p=0.25)	0.76 (p=0.076)
Other Party	0.49 (p=0.053)	-0.03 (p=0.8860)	0.52 (p=0.1412)
President			
Same Party	-0.54 (p=0.1929)	0.95 (p=0.0071)	1.50 (p=0.0048)
Other Party	0.39 (p=0.2949)	0.20 (p=0.4934)	0.19 (p=0.6925)
Governor			
Same Party	-0.58 (p=0.1349)	0.57 (p=0.0984)	1.15 (p=0.0268)
Other Party	0.07 (p=0.8528)	0.03 (p=0.9134)	0.03 (p=0.9413)
Leadership			
Same Party	-0.44 (p=0.2505)	0.73 (p=0.0568)	1.17 (p=0.0332)
Other Party	0.45 (p=0.2338)	-0.02 (p=0.9485)	0.47 (p=0.34493)

Appendix 4D: Treatment Effects by Party

In the next section of this appendix we consider whether effects differ by party, and whether effects differ by party strength. We consider party strength to ensure that our effects are not a function party salience. Since party salience is most likely to differ for weak partisans (Krupnikov and Piston 2015), we ensure that we still observe similar patterns for participants who identify as weak partisans.

Table 4D1: Weak Partisans, High Prejudice						
	Black Candidate, Same Party	White Candidate, Same Party	D-I-D	Black Candidate, Other Party	White Candidate, Other Party	D-I-D
Weak Partisans						
Δ Favorability	0.12 (p=0.82)	1.23 (p=0.03)	1.11 (p=0.15)	1.45 (p=0.005)	0.98 (p=0.0846)	-0.47 (p=0.53)
Δ Vote Willingness	-0.08 (p=0.83)	0.74 (p=0.05)	0.82 (p=0.12)	1.2 (p=0.002)	0.63 (p=0.12)	-0.57 (p=0.29)
Δ Gubernatorial Success	0.21 (p=0.69)	1.48 (p=0.002)	1.27 (p=0.07)	0.93 (p=0.06)	1.12 (p=0.05)	0.19 (p=0.80)
Δ Presidential Success	0.25 (p=0.66)	1.23 (p=0.02)	0.98 (p=0.19)	0.87 (p=0.03)	0.50 (p=0.41)	-0.37 (p=0.61)
Δ Leader	0.30 (p=0.59)	1.60 (p=0.002)	1.30 (p=0.07)	1.22 (p=0.004)	1.49 (p=0.01)	0.28 (p=0.69)

Appendix 4E: Treatment Effects by Different Correction Type

Since we have different types of corrections, we ensure that our effects are robust to both the unattributed and attributed corrections. Results for the attributed correction lone are shown in Table 4E1.

Table 4E1: Robustness Check: Effect of Attributed Correction on Candidate Evaluations, by Type of Candidate and Prejudice Level

	Black Candidate, Same Party	White Candidate, Same Party	Black Candidate, Other Party	White Candidate, Other Party
High Prejudice Participants:				
Δ Favorability	0.13 (p=0.82)	1.38 (p<.01)	1.94 (p<.01)	0.83 (p=.04)
Δ Vote Willingness	0	0.44 (p=0.19)	1.23 (p<.01)	0.53 (p=.10)
Δ Gubernatorial Success	0.32 (p=0.53)	1.24 (p<.01)	1.10 (p<.01)	1.04 (p=.01)
Δ Presidential Success	0.24 (p=0.66)	1.10 (p=.01)	1.05 (p=.01)	0.66 (p=.14)
Δ Leadership	0.42 (p=0.43)	1.10 (p=0.009)	1.52 (p<.01)	0.90 (p=0.055)
Low Prejudice Participants:				
Δ Favorability	0.84 (p=.008)	1.09 (p=.0018)	1.40 (p=0.0002)	-0.26 (p=.42)
Δ Vote Willingness	0.34 (p=0.11)	0.64 (p=.01)	0.40 (p=.18)	0.26 (p=.34)
Δ Gubernatorial Success	0.60 (p=0.03)	0.77 (p=.03)	1.08 (p=.01)	0.09 (p=.78)
Δ Presidential Success	0.54 (p=0.08)	0.84 (p=.02)	1.09 (p=.005)	0.19 (p=.58)
Δ Leadership	0.28 (p=0.35)	0.88 (p=0.02)	1.08 (p=0.007)	0.24 (p=0.51)

Note: Cells contain the effect of the correction (treatment – control) on evaluations. Positive (negative) scores mean that participants moved in a more positive (negative) direction. Favorability and likelihood of success as governor and president are measured on a 7-point scale. Vote willingness is measured on a 5-point scale. P-values come from two-sided tests comparing treatment (correction) and control (uncorrected) group means.

Table 4E2: Robustness Check: Effect of Unattributed Correction on Candidate Evaluations, by Type of Candidate and Prejudice Level

	Black Candidate, Same Party	White Candidate, Same Party	Black Candidate, Other Party	White Candidate, Other Party
High Prejudice Participants:				
Δ Favorability	0.22 (p=.64)	1.36 (p<.01)	0.62 (p=.07)	0.70 (p=.05)
Δ Vote Willingness	-0.29 (p=.43)	0.70 (p=.02)	0.66 (p=.01)	0.20 (p=.43)
Δ Gubernatorial Success	0.14 (p=.78)	1.37 (p<.01)	0.65 (p=.06)	0.61 (p=.08)
Δ Presidential Success	-0.28 (p=.56)	1.55 (p<.01)	0.65 (p=.03)	0.59 (p=.12)
Δ Leadership	0.05 (p=0.93)	1.54 (p<.01)	0.99 (p=0.004)	0.57 (p=0.13)
Low Prejudice Participants:				
Δ Favorability	0.71 (p=.03)	0.64 (p=.05)	0.88 (p=.005)	0.12 (p=.67)
Δ Vote Willingness	0.41 (p=.09)	0.36 (p=.13)	0.08 (p=.78)	0.08 (p=.75)
Δ Gubernatorial Success	0.60 (p=.05)	0.36 (p=.24)	0.77 (p=.02)	-0.002 (p=.96)
Δ Presidential Success	0.67 (p=.06)	0.40 (p=.22)	0.49 (p=.12)	-0.08 (p=.77)
Δ Leadership	0.56 (p=0.09)	0.41 (p=0.22)	0.95 (p=0.01)	0.08 (p=0.77)

Note: Cells contain the effect of the correction (treatment - control) on evaluations. Positive scores mean that participants moved in a more positive direction. Favorability and likelihood of success as governor and president are measured on a 7-point scale. Vote willingness is measured on a 5-point scales. P-values come from two-sided tests comparing treatment (correction) and control (uncorrected) group means.

Appendix 4F: Triple Interaction Models

4F1. Triple Interaction

The main analyses we present in text rely on t-tests and difference-in-difference comparisons. In this appendix we estimate a model that includes our treatments as a triple interaction between the partisanship of the participant relative to the candidate in the treatment, the race of the candidate and the correction. We present our results of a triple interaction which splits the data into participants with low racial prejudice and high racial prejudice in Tables 6A1 and 6A2. These results reinforce our previous conclusions.

Table 4F1: Triple Interaction Models (Low Prejudice Participants)

	DV = Favorability	DV = Vote Likelihood	DV = Effective Governor	DV = Effective President
Black (same party =0, correction =0)	-0.39 (0.33)	0.00 (0.00)	-0.36 (0.32)	-0.47 (0.34)
Same Party (Black =0, correction =0)	0.12 (0.28)	0.00 (0.00)	0.21 (0.28)	0.09 (0.29)
Correction (Black =0, same party = 0)	-0.01 (0.25)	0.00 (0.00)	-0.03 (0.24)	-0.12 (0.26)
Black*Same Party (correction =0)	0.45 (0.46)	0.00 (0.00)	0.46 (0.45)	0.49 (0.47)
Black*Correction (same party =0)	1.10*** (0.40)	0.00 (0.00)	0.94** (0.39)	0.89** (0.41)
Same Party*Correction (Black =0)	0.85** (0.36)	0.00 (0.00)	0.58* (0.35)	0.72* (0.37)
Black*Same Party*Correction	-1.17** (0.56)	0.00* (0.00)	-0.89 (0.55)	-0.88 (0.58)
Constant	3.21***	2.40***	3.39***	3.22***

	(0.20)	(0.16)	(0.19)	(0.20)
N	531	547	536	528

Note: Cell entries are OLS coefficients with standard errors in parentheses. Favorability and likelihood of success as governor and president are measured on a 7-point scale. Vote willingness is measured on a 5-point scale. Significance levels: ***p<.01, **p<.05, *p<.10.

Table 4F2: Triple Interaction Models (High Prejudice Participants)

	DV = Favorability	DV = Vote Likelihood	DV = Effective Governor	DV = Effective President
Black (same party =0, correction =0)	-0.12 (0.38)	-0.46 (0.30)	-0.07 (0.38)	-0.39 (0.39)
Same Party (Black =0, correction =0)	0.64 (0.34)	0.67** (0.27)	0.67** (0.34)	0.51 (0.35)
Correction (Black =0, same party =0)	0.74 (0.30)	0.35 (0.24)	0.81*** (0.31)	0.65** (0.31)
Black*Same Party (correction =0)	0.93 (0.55)	0.92** (0.43)	0.57 (0.54)	0.84 (0.56)
Black*Correction (same party =0)	0.33 (0.48)	0.50 (0.37)	0.00 (0.47)	0.14 (0.49)
Same Party*Correction	0.57 (0.44)	0.26 (0.34)	0.50 (0.43)	0.71 (0.44)
Black*Same Party*Correction	-1.56 (0.70)	-1.28** (0.56)	-1.08 (0.70)	-1.55** (0.72)
Constant	2.33 (0.25)	1.88*** (0.20)	2.50*** (0.25)	2.26*** (0.25)
N	277	286	276	273

Note: Cell entries are OLS coefficients with standard errors in parentheses. Favorability and likelihood of success as governor and president are measured on a 7-point scale. Vote willingness is measured on a 5-point scale.

Appendix 5: Additional Information on the Racial Prejudice Measure

As discussed in the main text, we used the standard four-question stereotype battery to measure racial prejudice (see Appendix A for question wordings). Following past research, we used these variables to create a racial prejudice score by subtracting participants' ratings of blacks from their ratings of whites (e.g., Kinder and Mendelberg 1995; Hutchings 2009; Piston 2010; Krupnikov and Piston 2015). We re-coded these variables so that the resulting scale runs from 0 (lowest prejudice) to 1 (highest prejudice). We then dichotomized the scale at the mid-point (0.5) and created an indicator variable, which takes the value of 1 if a participant scored above this value and a 0 otherwise.[1] This procedure was executed in R (R Core Team 2017).

To avoid priming racial considerations and contaminating the experimental treatments, we measured these prejudice variables at the very end of the survey. This approach is consistent with past research using similar measures. As Valentino et al. (2002: 78) note, "racial attitudes and other political predispositions are acquired at an early age and stable throughout the life span." Nonetheless, to ensure that our data are no different than previous studies that rely on racial attitudes, we test whether the treatments affected responses to the prejudice questions. We do so in several ways. In Appendix 5A we estimate models predicting racial attitudes by condition. In Appendix 5B we use a re-contact survey to ensure that racial attitudes are not affected by treatment assignment.

[1] We dichotomize at the scale mid-point (0.5) to increase comparability with other studies that use this same scale. In our data, the scale mid-point and the empirical median are both 0.5.

Appendix 5A: Models Predicting Racial Attitudes

We estimated five logistic regression models predicting prejudice level based on treatment assignment (see Table 5A1). In each model, the dependent variable is a binary indicator of a participant's prejudice classification (1=high prejudice, 0 otherwise), and the independent variable is a dummy variable for treatment assignment (1=correction, 0 otherwise). The results are broken down by type of candidate. As can be seen in the table, we find no evidence that our treatments affect measured levels of racial prejudice.

Table 5A1: Logit Models Predicting High Prejudice Classification

	Same Party, Black Candidate	Same Party, White Candidate	Other Party, Black Candidate	Other Party, White Candidate
Correction	-0.10	-0.19	-0.40	0.13
	(0.36)	(0.28)	(0.33)	(0.27)
Constant	-0.82***	-0.50**	-0.38	-0.63***
	(0.28)	(0.21)	(0.27)	(0.22)
N	161	236	180	256

Note: Cell entries are logit coefficients with standard errors in parentheses. Significance levels: ***p<.01, **p<.05, *p<.10.

We also assessed whether these estimates, which speak to the potential *average* effect of the treatments on prejudice, conceal individual-level heterogeneity. Put differently, it is possible that, although the average effect of the treatments on prejudice is insignificant, certain types of participants were in fact affected. Of particular interest here is education, as highly educated individuals are especially susceptible to social desirability bias (Karp and Brockington 2005). If our treatments led participants to respond to the prejudice questions in socially desirable ways (i.e., to express less animosity to blacks than they would otherwise), we would expect this effect to be concentrated among the highly educated. We assessed this possibility by estimating five additional logistic regression models (see Table 5A2). The dependent variable is again a

binary indicator of prejudice classification (1=high prejudice, 0=otherwise). The independent variables include the same dummy variable for treatment assignment (1=correction, 0 otherwise), an indicator for whether the participant has more than the median level of education (1=high education, 0 otherwise), and the interaction of these two variables. (The median participant in our sample had an Associate's degree; thus, "highly educated" participants are those who earned a Bachelor's degree or higher.)

As shown in the table, we find scant evidence that education affects our measure of racial prejudice even in conditions with a black politician. Out of ten coefficients involving education, only one approaches significance (p=.09, two-sided). Moreover, this occurs in the condition with the black candidate of the opposing party when no correction is offered. This means that in the uncorrected condition, people with higher levels of education may be slightly likely to give more positive evaluations – an outcome that is unlikely to produce the specific patterns of results observed in our analysis.

Table 5A2: Logit Models Predicting High Prejudice Classification (with Education)

	Same Party, Black Candidate	Same Party, White Candidate	Other Party, Black Candidate	Other Party, White Candidate
Correction	-0.34	0.14	-0.57	0.45
	(0.45)	(0.37)	(0.43)	(0.37)
Highly Educated	-0.83	-0.13	-0.94	0.20
	(0.65)	(0.42)	(0.56)	(0.44)
Correction*	0.82	-0.89	0.51	-0.73
Highly Educated	(0.80)	(0.58)	(0.69)	(0.55)
Constant	-0.61*	-0.44	-0.00	-0.72**
	(0.34)	(0.29)	(0.34)	(0.31)
N	159	235	180	255

Note: Cell entries are logit coefficients with standard errors in parentheses. Significance levels: ***p<.01, **p<.05, *p<.10.

In sum, we see little evidence that our conditions affected the prejudice measure. This is in line with previous findings demonstrating that racial attitudes are difficult to move (Valentino et al 2002).

Appendix 5B: Racial Attitudes Measured 2 Years Post-Treatment

Although our initial set of tests suggests that treatment did not affect racial attitudes, we conducted an additional study to reinforce this point. In this additional study we re-contacted participants who initially took our study in 2014. All people who participated in our initial study were eligible for this re-contact. We are able to conduct a re-contact study since our participants were recruited via the SSI panel. SSI maintains its panel by assigning unique panel ID numbers to each panel member.

Participants were recruited for the re-contact study in the same manner that they may be recruited for any other study that they take with SSI. There was no connection made between the re-contact study and the previous study in which they had participated almost two years ago. Moreover, given the passage of time, the consent form for the re-contact study actually came from a different institution than the original study, making the possibility that participants discovered any connection between the two unlikely.

In total, 312 participants took part in our re-contact study, which was expressly designed to measure racial attitudes (although, to decrease social desirability pressures, these measures were randomized within other batteries of personality measures that have no racial content) and susceptibility to social desirability pressures. Prior to analyzing these participants' racial attitudes we ensure that that there are no systematic effects that

led individuals to be more likely to participate in the re-contact (see Table 5B1). We find no systematic factors leading to participation in the re-contact.

Table 5B1: Logit Model Predicting Participation in Re-contact Survey

Condition 2	-0.21	-0.13
	(0.35)	(0.35)
Condition 3	-0.03	-0.01
	(0.35)	(0.35)
Condition 4	-0.26	-0.21
	(0.33)	(0.34)
Condition 5	0.19	0.26
	(0.38)	(0.39)
Condition 6	-0.27	-0.21
	(0.34)	(0.35)
Condition 7	-0.08	-0.01
	(0.35)	(0.36)
Condition 8	-0.32	-0.25
	(0.32)	(0.33)
Female		-0.24
		(0.15)
Education		-0.01
		(0.06)
Income		0.06
		(0.04)
Democrat		-0.09
		(0.18)
Republican		-0.23
		(0.20)
Constant	-0.69*	-0.80*
	(0.27)	(0.38)
N	1,031	1,031

Note: Cell entries are logit coefficients with standard errors in parentheses. Omitted reference group is Condition 1 (same party black candidate, no correction). Significance levels: ***p<.01, **p<.05, *p<.10

5B1. Susceptibility to Social Desirability

We focus first on social desirability issues. Here we rely on Berinsky and Levine's (2011) three-item social monitoring index. Self-monitoring is designed to measure the extent to which individuals are likely to change their behaviors to ensure that they behave in what they perceive to be a more socially desirable manner. As Weber et al. (2014) demonstrate, individuals who are higher in social desirability may have a higher tendency to misreport their racial attitudes. We measured participants' self-monitoring in the re-contact study.

We used these data to estimate the same five models presented above, using participants' social monitoring scores (instead of education) to measure the tendency to give socially desirable responses. Here we use racial attitude measured in the 2014 study. Since people who are high in self-monitoring are particularly susceptible to social desirability, if our conditions are affecting social desirability tendencies, then the inclusion of self-monitoring should highlight those effects. As shown in the table, we find no evidence that the experimental treatments are producing differential effects on reporting of prejudice even when we account for self-monitoring.

Table 5B2: Logit Models Predicting High Prejudice Classification (with Self Monitoring)

	Full Sample	Same Party, Black Candidate	Same Party, White Candidate	Other Party, Black Candidate	Other Party, White Candidate
Unattributed Correction	-0.08 (0.42)	1.57 (1.29)	-0.22 (0.91)	1.10 (1.16)	-0.13 (0.78)
Attributed Correction	-0.03 (0.42)	1.16 (1.35)	-0.97 (0.97)	-2.20 (1.41)	-0.07 (0.87)
High Self Monitor	0.06 (0.39)	0.69 (1.32)	0.17 (0.73)	-1.28 (1.30)	-0.13 (0.78)
Unattributed * High Self Monitor	0.15 (0.57)	-0.41 (1.65)	-0.64 (1.23)	1.90 (1.62)	0.24 (1.12)
Attributed * High Self Monitor	-0.12 (0.62)	-0.47 (1.79)	-0.02 (1.36)	-16.18 (1769.26)	0.54 (1.19)
Constant	-0.61** (0.29)	-2.08** (1.06)	-0.29 (0.54)	1.10 (1.16)	-0.15 (0.56)
N	296	49	67	57	72

Note: Cell entries are logit coefficients with standard errors in parentheses. Significance levels: ***p<.01, **p<.05, *p<.10

5B2. Re-Measurement of Racial Attitudes

During the re-contact survey we also re-measured participants' racial attitudes using the same four-question index originally used in 2014. We use these data to conduct two additional tests: (1) whether participants expressed similar levels of racial prejudice at both points in time, which would suggest that our treatments did not affect their initial responses; and (2) whether the main in-text results are robust to using the prejudice scale from the re-contact data.

We begin with the relationship between participants' initial responses to the prejudice questions and their responses at the time of re-contact. We present the

correlations and corresponding p-values for all four measures in Table 5B3. We find strong positive over-time correlation for each of the four questions (p<.001 for all), suggesting that the initial prejudice measures captured relatively stable racial attitudes.

Table 5B3: Correlation Between Prejudice Measures at Time of Initial Contact and at Re-Contact

Question	Correlation
Whites hard-working	r=0.38 (p<.001)
Blacks hard-working	r=0.56 (p<.001)
Whites intelligent	r=0.46 (p<.001)
Blacks intelligent	r=0.50 (p<.001)

Table 5B4: Correlation Between Racial Prejudice Measures, ANES

	Whites
2000-2002	0.35
2002-2004	0.32
2000-2004	0.38

Cell entries are Pearson's correlation coefficients between warmth (0-100 feeling thermometer) toward the relevant group in one year and in another. Source: Piston (2018).

Moreover, our correlations are in-line with ANES over-time panel correlations that are typically used to demonstrate the stability of racial attitudes (see Table 5B4). Even more importantly, when we see shifts in racial attitudes these shifts are not systematic to treatment assignment in 2014. Had there been any systematic factors that led to changes being more likely if participants were initially in one particular condition, this would have suggested that the conditions had differential effects on prejudice measures. The absence of systematic effects points to the idea that condition assignment did not lead systematic differences in racial attitude measures.

Next, we consider whether our main patterns are robust to using the prejudice measures from the re-contact. It is important to note that these tests have significantly lower power than those presented in the main text (the re-contact N=312, reducing our

sample size by nearly three times). Although significance levels change given the lower power, the key pattern is evident on all four measured dependent variables. That is, on all dependent variables, the correction of misinformation about a same party black candidate has a larger effect among low prejudice participants than among high prejudice participants.

Table 5B5: Effect of Correction on Candidate Evaluations, by Type of Candidate and Prejudice Level

	Black Candidate, Same Party	White Candidate, Same Party	Black Candidate, Other Party	White Candidate, Other Party
High Prejudice Participants:				
Δ Favorability	0.50 (p=.45)	1.39 (p<.01)	1.39 (p<.01)	1.32 (p<.01)
Δ Vote Willingness	0.10 (p=.84)	0.38 (p=.31)	0.47 (p=.29)	0.66 (p=.08)
Δ Gubernatorial Success	0.33 (p=.61)	1.25 (p<.01)	1.39 (p<.01)	1.00 (p=.08)
Δ Presidential Success	0.40 (p=.58)	1.08 (p=.02)	0.99 (p=.01)	0.91 (p=.08)
Low Prejudice Participants:				
Δ Favorability	0.81 (p=.12)	1.32 (p<.01)	1.78 (p<.01)	0.79 (p=.14)
Δ Vote Willingness	0.46 (p=.22)	0.26 (p=.47)	0.66 (p=.23)	-0.02 (p=.98)
Δ Gubernatorial Success	0.39 (p=.42)	0.58 (p=.20)	1.66 (p<.01)	0.76 (p=.17)
Δ Presidential Success	0.56 (p=.30)	0.90 (p=.05)	1.58 (p<.01)	0.42 (p=.49)

Note: Cells contain the effect of the correction (treatment - control) on evaluations. Positive (negative) scores mean that participants moved in a more positive (negative) direction. Favorability and likelihood of success as governor and president are measured on a 7-point scale. Vote willingness is measured on a 5-point scale. P-values come from two-sided tests comparing treatment (correction) and control (uncorrected) group means.

Appendix 7: Comparison of Candidate Suits

In order to ensure that our results are not a function of differences in individual response to candidate suits and the use of an American flag pin, we conducted a secondary check. In this check we recruited a group of 198 participants from Amazon's Mechanical Turk. We randomly assigned these participants to a photo of a candidate with the flag pin or to a photo of a candidate without the flag. We also retained identical text to the baseline condition. Post-treatment we asked participants to select which term best describes the politician from a set of terms (in randomized order).

The results are presented in Table 7A1 below and suggest that there are few differences between the candidate with and without the flag pin. In particular, we find no difference on the selection of the term "patriotic" – a characterization most likely to be affected by the flag pin.

Table 7A1: No Difference in Evaluations of Candidates with and without Flag Pins

	All Participants			White Participants		
	Candidate with flag pin (N=98)	Candidate without flag pin (N=100)	Difference p-value	Candidate with flag pin (N=83)	Candidate without flag pin (N=84)	Difference p-value
Patriotic	14.3%	14.%	p=0.9543	13.3%	14.3%	p=0.8476
Honest	8.2%	7%	p=0.7586	8.4%	6.0%	p=0.5376
Old	40.8%	41%	p=0.9792	43.4%	39.3%	p=0.5943
Kind	8.2%	13%	p=0.2714	8.4%	13.1%	p=0.3344
Friendly	24.5%	30%	p=0.3866	25.3%	29.8%	p=0.5217

Online Appendix References

Berinsky, Adam J. and Howard G. Lavine. 2011. "Self-Monitoring and Political Attitudes." In John Aldrich and Kathleen M. McGraw (ed.), *Improving Public Opinion Surveys: Interdisciplinary Innovation and the American National Election Studies*. Princeton: Princeton University Press.

Druckman, James N, Erik Peterson, and Rune Slothuus. 2013. "How Elite Polarization Affects Public Opinion Formation." *American Political Science Review* 107(1) 57-79.

Hutchings, Vincent L. 2009. "Change or More of the Same: Evaluating Racial Policy Preferences in the Obama Era." *Public Opinion Quarterly* 73(5): 917-942.

Jerit, Jennifer, Jason Barabas, and Scott Clifford. 2013. "Comparing Contemporaneous Laboratory and Field Experiments on Media Effects." *Public Opinion Quarterly* 77(1) 256-282.

Kam, Cindy D. and Franzese, Robert J. Jr. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis*. Ann Arbor, MI: University of Michigan Press.

Karp, Jeffrey A. and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *Journal of Politics* 67(3): 825-840.

Kinder, Donald R. and Tali Mendelberg. 1995. "Cracks in American Apartheid: The Political Impact of Prejudice among Desegregated Whites." *Journal of Politics* 57(2): 402-424.

Krupnikov, Yanna and Spencer Piston. 2015. "Racial Prejudice, Partisanship and White Turnout in Elections with Black Candidates." *Political Behavior* 37(2): 397-418.

Piston, Spencer. 2010. "How Explicit Racial Prejudice Hurt Obama in the 2008 Election." *Political Behavior* 32(4): 431-451.

Piston, Spencer. 2018. *Class Attitudes in America*. New York: Cambridge University Press.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Tesler, Michael. 2013. "The Return of Old-Fashioned Racism to White Americans' Partisan Preferences in the Early Obama Era." *Journal of Politics* 75(1): 110-123.

Valentino, Nicholas, Vincent Hutchings, and Ismail White. 2002. "Cues that Matter: How Political Ads Prime Racial Attitudes During Campaigns." *American Political Science Review* 96(1): 75-90.

Weber, Christopher R., Howard Lavine, Leonie Huddy, and Christopher M. Federico. 2014. "Placing Racial Stereotypes in Context: Social Desirability and the Politics of Racial Hostility." *American Journal of Political Science* 58(1): 63-78.