Abstract:

Attempts to correct political misperceptions often fail. The dominant theoretical explanation for this failure comes from psychological research on motivated reasoning. We identify a novel source of motivated reasoning in response to corrective information: the justification of socially undesirable preferences. Further, we demonstrate that this motivation can, under certain conditions, overpower the motivation to maintain congruence. Our empirical test is a national survey experiment that asks participants to reconcile partisan motivations and the motivation to justify voting against a racial minority candidate. Consistent with our argument, racially prejudiced participants dismiss corrections when misinformation is essential to justify voting against a black candidate of their own party, but accept corrections about an otherwise identical candidate of the opposing party. These results provide new insight into the persistence of certain forms of political misinformation.

Word Count: 3,388

* Corresponding author. We thank Jamie Druckman, Seth Frey, Kevin Mullinix, anonymous reviewers, and participants at the 2014 annual meeting of the American Political Science Association for helpful comments on earlier versions. The authors declare no conflicts of interest. The data, code, and any additional materials required to replicate all analyses in this article are available at the Journal of Experimental Political Science Dataverse within the Harvard Dataverse Network, at: doi:1010.7910/DVN/7TRL0I.
Many citizens hold misperceptions about politically relevant facts.\footnote{Following Nyhan and Reifler (2010: 305), we define misperceptions as “cases in which people’s beliefs about factual matters are not supported by clear evidence and expert opinion” (cf., Gaines et al. 2007).} Attempts to change these false beliefs by providing correct information often fail (e.g., Nyhan and Reifler 2010). Why do misperceptions persist, even in the face of corrections?\footnote{We refer to communications that attempt to correct misperceptions as “corrective messages” or “corrections.”} The dominant theoretical explanation for this persistence comes from research on motivated reasoning (Flynn et al. 2017), or people’s tendency to accept information that is congruent with existing beliefs and to dismiss incongruent information (Kunda 1990, Taber and Lodge 2006). According to this account, people are motivated to accept (dismiss) corrections that are congruent (incongruent) with existing partisan or ideological preferences.

Although the motivation to maintain congruence is important, people’s responsiveness to new information is often the product of numerous — at times competing — motivations (Druckman 2012). In this article, we identify a novel source of motivated reasoning in response to corrective information: the justification of socially undesirable preferences. When people exhibit preferences or behaviors that may be perceived as socially undesirable, they prefer to have information that justifies those preferences and behavior (Duff et al. 2007). Indeed, psychological research on variants of motivated reasoning indicates that people are generally accepting of information that allows them to justify pre-determined conclusions (Ditto and Lopez 1992). In contrast, people are more dismissive of information that does not allow for this type of justification (Ditto and Lopez 1992).
Under certain conditions, this justification motivation may lead to different outcomes than congruence-based motivated reasoning. Research demonstrates that people who want to justify disliking person whom they, following social standards, should like are often motivated to dismiss any positive information about this person and cling to negative information (Ruscher et al. 2000). On the other hand, people who are already armed with a (socially appropriate) reason for disliking another person are often less motivated to dismiss new incongruent information (Ruscher et al. 2000). Applying these arguments to the case of misinformation, people should be least responsive to corrections (i.e., more inclined to retain misinformation) when misinformation can be used to justify an otherwise socially undesirable preference.

There are many contexts in which a person may need to justify — even to themselves — a socially undesirable preference, but the particular case we focus on is an unwillingness to support a black candidate. Race a critical factor in voter decision-making (Hutchings and Piston 2011, Piston 2010); moreover, people are hesitant to appear in any way racially prejudiced (Sigelman et al. 1995). As a result, people actively search for justifications for voting against black candidates (Krupnikov et al. 2016; Kline and Stout 2015).

Our argument is captured by the following example. Consider hypothetical Voter A, who does not want to vote for his party’s candidate, Candidate A, because Candidate A is black. During an election campaign Voter A receives — and believes — some negative information (X) about Candidate A. X provides Voter A with a justification for opposing Candidate A. In particular, X allows Voter A to justify opposing Candidate A by providing a reason that has nothing to do with Candidate A’s race, making X very
useful for this voter. Later in the campaign credible evidence emerges that X is false. Yet, since X provides an important justification for opposing Candidate A, Voter A is motivated to dismiss the correction and retain the misinformation about Candidate A. As a result, the correction is unlikely to affect the voter’s evaluations of Candidate A.

Now consider hypothetical Voter B. Voter B is of a different party than Candidate A. Voter B receives — and believes — the same negative information (X) about Candidate A. Voter B also later receives a correction — the same correction that Voter A received — indicating that X is incorrect. Voter B also does not want to vote for Candidate A; however, Voter B’s justification rests on partisanship and, potentially, many other pieces of information (e.g., issue positions). In this case, Voter B may be responsive to the correction of X, as X is relatively unnecessary for justifying opposition. Indeed, since Candidate A is of a different party and Voter B needs no additional justification for opposing the candidate, the voter may even respond more favorably to the candidate after the correction.

This example is not to suggest that people are unmotivated to retain negative information about opposing party candidates. Indeed, people will generally cling to information that reinforces existing preferences. Moreover, this is also not to suggest that members of the black candidate’s own party should in all cases be more likely than members of the opposing party to retain misinformation. Rather, we suggest that there are conditions under which people may be somewhat more motivated to resist corrections of misinformation about members of their own party. In lower salience, non-presidential elections — the types of elections which are more likely to include black candidates — can be more difficult to justify opposing a co-partisan candidate, which means every
negative piece of information about that candidate provides justification.

We test the observable implications of our argument for candidate evaluations using a national survey experiment that varies two of the most salient characteristics of political candidates: partisan affiliation and race (Krupnikov and Piston 2015). We focus on the extent to which candidate evaluations change in response to corrections (i.e., more positive evaluations would signal responsiveness to the correction). Consistent with our argument, we demonstrate that people are least responsive to messages that seek to correct misinformation that justifies what may be an otherwise socially undesirable preference: opposing a black candidate of their own party.

The Experiment

Design

We asked our participants — a nationally representative sample of adults (N=1,031) — to read a news-brief about a politician. While the text and layout of the news-brief remained identical for all participants, we randomized the type of politician being described and whether a correction was presented. Participants were randomly assigned into one of 8 conditions: 4 (type of politician: same party black, same party white, other party black, other party white) x 2 (misinformation: uncorrected, corrected) (see Table 1).  

---

3 We used Survey Sampling International (SSI) to collect the data. SSI recruited a sample that matches the national population on key benchmarks. See Appendix 3 for sample demographics.
4 Randomization checks confirm that conditions were balanced on pre-treatment covariates (see Appendix 3).
5 We found no evidence of heterogeneous treatment effects by party (see Appendix 4). Although significance levels change due to lower power in the split samples, the key pattern is apparent among both Democrats and Republicans, as well as weak partisans.
The misinformation concerned allegations that the legislator described in the news-brief used his position on a key committee to secure a tax loophole for a campaign donor, citing contribution records as evidence (see Appendix 2 for full text of all treatments). This information was designed to reflect negatively on the legislator. The correction cited state legislative records, which clearly showed that the legislator in question did not actually secure the loophole; in fact, a different member of the committee introduced the amendment and worked to secure its passage. The correction also indicated that the recipient of the tax break had not actually donated to the legislator’s campaign. Race was cued using photographs, with both images and name pre-tested to rule out confounding effects. A full list of experimental conditions is displayed in Table 1.

Table 1: Experimental Conditions

<table>
<thead>
<tr>
<th>No Correction</th>
<th>Black Candidate, Same Party</th>
<th>White Candidate, Same Party</th>
<th>Black Candidate, Other Party</th>
<th>White Candidate, Other Party</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 5</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 2</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 3</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 4</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 5</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 6</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 7</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
<tr>
<td>Group 8</td>
<td>Black Candidate, Same Party</td>
<td>White Candidate, Same Party</td>
<td>Black Candidate, Other Party</td>
<td>White Candidate, Other Party</td>
</tr>
</tbody>
</table>

---

6 Previous research indicates that scandals involving professional responsibilities are particularly deleterious (Doherty, Dowling, and Miller 2011); thus, our design constitutes a tough test for the effectiveness of corrections.

7 To assess the robustness of the correction effect, we randomized the source of the correction. Some participants (n=264) read a correction attributed to “PolitiCheck,” a fictional group described as “a non-partisan fact-checking organization whose goal is to adjudicate factual disputes.” Other participants (n=383) read an identical correction that was not attributed to any source. The key results are substantively the same across the attributed and unattributed corrections (see Appendix 4). Thus, we merge data from both treatment groups.

8 Attractiveness of candidate was pretested using an earlier sample (N=293) from Amazon.com’s Mechanical Turk (Berinsky et al. 2012). We find no significant differences in ratings: the white candidate was rated 2.71 and the black candidate 2.70 on a 5-point attractiveness scale. A second MTurk sample (N=198) also shows no difference in response to candidates based on their suits (see Appendix 7).
Measures

We next asked our participants a series of questions to measure the extent to which they viewed the politician from the news-brief favorably (see Appendix 2 for all questions). These included measures of overall favorability, likelihood of voting for the politician in a future election, likelihood that the politician would make an effective governor or president in the future, and the candidate’s leadership capabilities. After measuring these variables, we gauged participants’ racial attitudes using the standard racial stereotype battery, which consists of four questions. Following previous research, we use 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 dichotomize the resulting scale at the median and refer to people who scored above the median prejudice level as having “negative racial attitudes” (Piston 2010).

Observable Implications

Given this design, we expect that, all things being equal, individuals with negative racial attitudes will prefer the white politician to the black politician (Hutchings and Piston 2011). As a result, these participants with negative racial attitudes will allow us to cleanly test for justification-based motivated reasoning in response to corrections.

People with negative racial attitudes often dislike black candidates (Ehrlinger et

---

9 To avoid priming racial considerations, which could contaminate the experimental treatments, these questions were asked at the very end of the survey. We conducted a series of tests to ensure that the treatments did not affect participants’ answers to the prejudice items (see Appendix 5 for more detail). Across all these tests, we find no evidence that treatment assignment predicts measured levels of racial prejudice. This measurement approach is consistent with previous research (Valentino et al 2002).
al. 2011; Highton 2004; Hutchings 2009; Lewis-Beck, Tien, and Nadeau 2010; Piston 2010; Redlawsk, Tolbert, and Franko 2010; Schaffner 2011; Tesler and Sears 2010); however, it is socially undesirable to acknowledge that one would not vote for a candidate due solely to the candidate’s race (Berinsky 2004). In particular, looking unfavorably at a black candidate of one’s own party may seem like a conspicuous act of prejudice (Krupnikov and Piston 2015). Faced with a black candidate of their own party, negative — but not explicitly racial — misinformation about the black politician proves useful: it offers people with negative racial attitudes a convenient, non-racial explanation for their lack of support.

In contrast, a black politician of the opposing party presents a different scenario for people with negative racial attitudes. In this case, expressing opposition is expected on the basis of partisanship alone and thus unlikely to be seen as overtly racial. In this case, misinformation is no longer pivotal, as party offers sufficient justification for opposing the black candidate. Thus, there is no justification-based motivation to retain misinformation, and corrections should prove effective, even for a person with negative racial attitudes.

Our experimental design allows us to test the observable implications of our predictions. We focus on the extent to which participants’ perceptions of the candidates change in response to corrections, rather than on the evaluation of the correction itself. As a result, our key prediction is that corrections should prove least effective in changing candidate evaluations when participants with negative racial attitudes are faced with a black candidate of their own party. In contrast, these same types of participants should be

---

10 One limitation with this design is that we cannot identify the underlying cognitive mechanisms.
more likely to change their evaluations of a candidate in response to a correction about a black candidate of the opposing party.\textsuperscript{11}

In what follows, we present our results as group means. We focus our in-text discussion on the favorability outcome variable, which is arguably most proximate to the treatments; we present results using all included outcome measures in the appendix. We also present additional empirical tests of our results and alternative explanations of our findings. We conclude with a discussion of limitations and implications of our results.

\textbf{Results}

Our estimand of interest is the difference in favorability ratings between participants who saw a correction and participants who did not. Following previous research, we use a statistically significant shift as proxy for correction effectiveness, and lack of a significant difference as a proxy for ineffectiveness and the continued influence of misinformation (Fridkin, Kenney, and Wintersieck 2015). A significant, positive (negative) difference means that views of the politician became more (less) favorable after reading the correction. The absence of group differences would suggest that favorability was the same pre- and post-correction. We consider each of these differences within candidate type. We analyze participants with negative racial attitudes separately from those with more positive or neutral racial attitudes, as the former group offers the clearest test of our argument.

We display our results graphically in Figure 1, which gives mean favorability ratings in each experimental condition, among participants with positive/neutral (left-}

\textsuperscript{11} Because our design randomizes both candidate type and correction, we are able to straightforwardly test whether the effect of the correction depends on the type of candidate under consideration (Kam and Trussler 2017: 794).
hand panel) or negative (right-hand panel) racial attitudes. The dark gray bars display mean ratings among participants who did not read a correction, while the light gray bars display ratings among participants who did read a correction. The size and significance of this “correction effect” — and relevant difference-in-difference estimates — are presented in Table 2 (see Appendix 4 for analysis of the other outcome measures).

We first examine results among participants with negative racial attitudes. Among these participants, we expect to observe the smallest group differences — meaning that the correction was least effective — when these participants are confronted with a black candidate of their own party.

Looking at the left-most pair of bars in the right panel of Figure 1, we see the predicted pattern of results: the correction had no effect when participants with negative racial attitudes were confronted with a black candidate of their own party. Participants who receive the correction and those who do not rated the black politician similarly (difference=0.17, p=0.67).\(^\text{12}\) This lack of a significant correction effect is shown graphically in the left-most pair of bars in the right-hand panel of Figure 1. In contrast, as shown in the second pair of bars in the right-hand panel of Figure 1, when the politician is a white co-partisan, we see that the correction was largely effective. Participants who receive the correction rate the white politician significantly and substantively more positively than those who did not receive the correction (difference=1.35, p<0.01). As shown in Table 2, difference-in-difference analyses confirm that, among participants with negative racial attitudes, the correction effect is significantly different for white and black

\(^{12}\) All reported p-values come from two-sided tests.
We observe similar patterns across the other outcome measures (see Appendix 4).

### Table 2: Mean Favorability Ratings of Same-Party Candidates among Low and High Prejudice Participants

<table>
<thead>
<tr>
<th></th>
<th>Low Prejudice Participants</th>
<th>High Prejudice Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Correction</td>
<td>Correction</td>
</tr>
<tr>
<td>Same Party, Black</td>
<td>3.38 (1.35)</td>
<td>4.16 (1.33)</td>
</tr>
<tr>
<td>Same Party, White</td>
<td>3.32 (1.79)</td>
<td>4.16 (1.44)</td>
</tr>
<tr>
<td>D-I-D</td>
<td><strong>0.06</strong> (p=0.87)</td>
<td>D-I-D</td>
</tr>
</tbody>
</table>

Note: Left panel contains results among low prejudice sample (N=683); right panel contains results among high prejudice sample (N=333). “No Correction” and “Correction” columns contain mean favorability ratings for same party black (row 3) or same party white (row 4) candidates (standard deviations in parentheses). “Difference” columns contain the effect of the correction on favorability ratings for given type of candidate (two-sided p-value in parentheses). D-I-D is the difference-in-difference estimate comparing the effect of the correction across black and white candidates (two-sided p-value in parentheses).

Although the participants with negative racial attitudes in the baseline conditions do initially rate the black candidate somewhat more positively than the white candidate, our results are not a function of ceiling effects. Across our outcome measures, participants with negative racial attitudes rate the black candidate of their own party just below the mean of a 7-point scale at baseline, which leaves open the entire positive end of the scale (i.e., values 5, 6 and 7). Even more importantly, post-correction the participants with negative racial attitudes rate the white candidate more positively than the black candidate of their own party. This suggests that these participants are willing to

---

13 We verified these results using other outcome measures and in models that include triple interactions (see Appendix 4). Results reinforce the patterns demonstrated in Figure 1 and Table 2.
go to higher ends of the rating scale — but not for the black candidate of their own party (see Appendix 4). In short, across all of our outcomes measures, while the correction eliminated the deleterious effects of misinformation for the white co-partisan, it failed to do so for the black co-partisan.

Figure 1: Candidate Favorability Ratings by Experimental Condition and Prejudice Level

Note: Brackets contain 95% confidence intervals. Left-hand panel includes low prejudice participants; right-hand panel includes high prejudice participants.

Although the correction proves ineffective for a black co-partisan, it proves more effective for a black politician of the opposing party (see third pair of bars in the right-hand panel of Figure 1). Here we see that the correction leads people to rate the black politician significantly more favorably (difference=1.07, p<.01). The correction is also effective for the white politician of the opposing party, also increasing his favorability rating (difference=0.77, p=0.02). As we might expect, however, this increase in favorability is somewhat less than the favorability increase for the white politician of the participants’ own party. These results are consistent with our argument that some forms
of misinformation are more pivotal than other forms. For voters with negative racial attitudes, misinformation is most pivotal when they are faced with a black politician of their own party.

Further, results change when we consider participants with more positive or neutral racial attitudes. Turning to these participants, we see that they are equally responsive to corrections about their own party’s politician, regardless of the politician’s race (black politician: difference=0.78, p<0.01; white politician: difference=0.84, p<0.01). Moreover, these participants are more responsive to corrections about the black politician of the opposing party than about the white politician of the opposing party. Although these participants do rate both politicians more favorably after receiving the correction, the effect is significantly more pronounced for the black candidate, which is in line with research indicating that people with positive racial attitudes often prefer black politicians to white politicians (Tesler and Sears 2010).

**Social Desirability and Robustness**

We take several steps to consider whether our results could be due to social desirability pressures playing differential roles in conditions, affecting our measures of racial attitudes, and thereby producing the patterns in this study. Although the fact that the study is conducted online does limit social desirability pressures, to address social desirability more fully we conducted a re-contact study of our participants. This study took place two years after the completion of the original study. We re-measured participants’ racial attitudes and measured their tendency to change behavior in response to social desirability pressures (Berinsky and Lavine 2012). Of our original sample, 312 participants took part in the re-contact (details are provided in Appendix 5). The re-
contact study demonstrated a high degree of consistency in participants’ racial attitudes, suggesting that the racial attitudes measured in our original study were unlikely a function of our particular treatments. Moreover, the re-contact showed no systematic response differences among individuals more susceptible to socially desirable responding (Berinsky and Lavine 2012). Finally, we replicate our key patterns using racial attitudes measured in the re-contact study. Collectively, although it is impossible to rule out social desirability fully, these additional tests work to weaken social desirability as an alternative explanation for our results.

**Discussion and Conclusion**

Many citizens hold misperceptions about political facts and engage in motivated reasoning when presented with corrections. Our goal here was to highlight a previously unidentified source of motivation: the desire to justify socially undesirable preferences. As we show, participants with negative racial attitudes are much more likely to retain misinformation about black co-partisans than black candidates of the opposing party. This is because, in this case, the misinformation offers a pivotal, *non-racial* justification for opposing a co-partisan candidate.

Before considering the implications of this research, we consider potential limitations. First, it is important to note that our study focuses on a non-racial piece of misinformation (i.e., special interest influence). It is possible that our results would change if the misinformation were racially tinged. That said, it is unlikely that explicitly racial misinformation will be viewed as pivotal by people who acknowledge the socially undesirability of opposing black candidates. Thus, our focus on non-racial information is
more reflective of the type of information people are likely to employ as justifications for opposing co-partisan minority candidates.

Second, our study represents a low-information election. Our participants made their judgments based on race, partisanship, and the information provided in our treatments. In other contexts (e.g., presidential elections), people have much more information with which to justify socially undesirable choices. Another possibility is that, in high profile elections, opposing partisans may want to justify a particularly vitriolic dislike for a black candidate, and as a result be more motivated to retain misinformation. Further, we note that many — if not most — elections are low information elections, and most black candidates run for lower offices (Krupnikov and Piston 2015). As a result, our study reflects the informational conditions under which people are most likely to encounter black candidates.

Finally, our study focuses on the observable outcomes of misinformation retention. Here we proposed a theory that rests on the intersection of motivated reasoning and racial attitudes, which, in turn, suggests the types of changes in candidate evaluations that should be expected under various conditions. Although our experimental results are consistent with our predictions, we acknowledge that the underlying psychological process linking misinformation to evaluations is beyond the grasp of this particular experimental design.

Nonetheless, the patterns reported here have implications for the way we view misinformation in contemporary politics. Citizens, these results suggest, may not be powerless in the face of their own biases. Rather, our results suggest that when they are
seeking to justify socially undesirable preferences, people may dismiss corrections for a much simpler reason: misinformation is convenient.
References


Schaffner, Brian F. 2011. “Racial Salience and the Obama Vote.” Political Psychology


