

Follow the Racist?

The Consequences of Trump's Expressions of Prejudice for Mass Rhetoric

Brian F. Schaffner
Newhouse Professor of Civic Studies
Department of Political Science & Tisch College
Tufts University
brian.schaffner@tufts.edu
Packard Hall 006
Medford, MA 02155

Abstract

Much of the literature on public opinion and political communication posits that elites are influential in how people think and talk about politics. Ideally, elite behavior provides a normatively desirable model for citizens to follow. But what happens when a prominent political elite violates norms of political and social discourse? The 2016 presidential campaign provided an example of such a dynamic, featuring a presidential candidate who made frequent comments that were explicitly prejudiced. What consequence did Trump's campaign have on the nature of mass rhetoric regarding issues of race? To answer this question, I use data from experiments fielded on the 2016 and 2017 Cooperative Congressional Election Studies. These experiments show evidence of a so-called "Trump Effect." That is, exposure to Trump's prejudiced statements made people more more likely to write offensive things, not only about the groups targeted by Trump, but sometimes about other identity groups as well. By contrast, in the second experiment, exposure to a quotation from Clinton lamenting the prejudice of Trump's supporters made people less likely to say offensive things about the target group.

Much of the literature on public opinion and political communication posits that elites are influential in affecting how people think about politics (Zaller, 1992). For the most part, this research has examined how elite cues or frames can affect discrete political choices, such as the positions that citizens take on issues or the candidates voters choose in elections. However, some scholars have also examined the extent to which elite rhetoric influences how people talk about politics. For example, individuals who are exposed to uncivil elite discourse are more likely to be uncivil in their own discussions about politics (Gervais, 2015, 2014). But how far does the influence of elite rhetoric extend?

In this paper, I examine whether exposure to prejudiced rhetoric from a prominent politician – Republican presidential nominee Donald Trump – affects what respondents say about identity groups. During the 2016 campaign, Costello (2016) posited the existence of a “Trump Effect” – the notion that Trump’s offensive rhetoric on the campaign trail was affecting the way that the mass public was talking about racial and ethnic minorities. Their survey found that “Teachers report that students have been ‘emboldened’ to use slurs, engage in name-calling and make inflammatory statements toward each other” (p. 10). On one hand, the notion that politicians could have such an effect on expressions of prejudice by members of the public is a remarkable one. After all, the research on racial rhetoric in American politics has long held that a “norm of racial equality” leads Americans to reject elite rhetoric that is explicitly racist (Mendelberg, 2001). If the public rejects appeals that are based on explicit racism, then one would also expect that public rhetoric would be unmoved by racist statements made by a prominent politician. On the other hand, recent scholarship suggests that the norm of racial equality has weakened in recent years and that white Americans no longer reject explicit racism as much as they once did (Valentino, Neuner, and Vandebroek, 2018). If this is true, then we might expect that the rhetoric of some white Americans could also be influenced by how politicians talk about racial and ethnic minorities.

Indeed, rather than simply reject Trump’s rhetoric, many citizens may have instead taken it as a signal about norms related to the expression of prejudice. Since Allport’s (1954) classic work on prejudice, scholars have understood that one’s adherence to and expression of prejudicial views is influenced by their desire to conform to the norms of their social situations. Of course, norms are often disputed or unclear, leading individuals to search for cues about what is or is not expected. In this paper, I report findings from two experiments designed to test whether Trump’s prejudiced rhetoric influences how individuals speak about other identity groups. My findings support the theory that Trump’s remarks serve as a cue regarding the norms of what constitutes acceptable speech. Specifically, I find that whites do react to his rhetoric by saying more negative and offensive things not only about the group targeted by Trump’s remarks, but often by commenting more negatively about other groups as well.

Norms, Signaling, and Expressions of Prejudice

Social conformity has long been understood to play an important role not only in the extent to which individuals hold prejudicial attitudes, but also on the extent to which they express those attitudes to others (Allport, 1954). Internally held prejudice is generally difficult to change, at least in the short term (Paluck and Green, 2009). By contrast, peoples’ willingness (or reluctance) to *express* prejudicial attitudes outwardly does appear to be more sensitive to their understandings of societal or group norms (Blanchard et al., 1994; Blanchard, Lilly, and Vaughn, 1991; Zitek and Hebl, 2007; Paluck, 2009; Crandall and Eshleman, 2003). In other words, people may not easily give up their prejudicial viewpoints, but they do appear to react to signals about norms regarding whether it is ok to express such views publicly.

For the most part, there has been a clear social norm in the U.S. regarding the expres-

sion of prejudice toward people of other races and ethnicities for many decades (Dovidio and Gaertner, 1986; Rokeach and Ball-Rokeach, 1989). This is evident from the fact that expressions of prejudice toward racial and ethnic minority groups dropped sharply since the middle of the twentieth century even though unobtrusive measures reveal that actual prejudice has not declined nearly as much (Crosby, Bromley, and Saxe, 1980; Hurwitz and Peffley, 1992; Kuklinski, Cobb, and Gilens, 1997). Thus, it seems that at least part of the reason that people express less prejudice toward minority racial groups is to conform to social norms that discourage such speech.

Of course, it is natural to assume that citizens have differing conceptions of what is or is not acceptable when it comes to expressing prejudice. It is also reasonable to imagine that some people are more confident about what these norms are compared to others. Indeed, this variance and uncertainty reflects the fact that norms are often ambiguous and vary significantly across different situations and contexts. Accordingly, people rely on cues from others regarding what the norms are in any particular situation (Cialdini and Trost, 1998). Scholars have demonstrated this cue-taking behavior with a variety of novel experiments. For example, Blanchard et al. (1994) allowed research subjects to overhear another subject's responses to questions related to racism before they themselves answered the same questions. However, the ostensibly overheard responses were actually given by a confederate whose remarks were randomly assigned. When the subject overheard remarks from the confederate condoning racism, that subject was also more likely to condone it, and when a subject overheard the confederate condemning racism, he was more likely to also express condemnation. This finding has been replicated by several similar studies focusing on a variety of different types of prejudice (Blanchard, Lilly, and Vaughn, 1991; Zitek and Hebl, 2007).

Cues can also come from sources beyond an individual's peer network. For example, Paluck (2009) finds that radio programs are effective in providing signals about norms related to prejudice. The study involved randomizing the types of radio programs that individuals

in Rwanda were exposed to, with an aim toward examining the impact of a program that focused on themes of ethnic reconciliation. Notably, being exposed to the anti-prejudice program did not affect listeners' actual beliefs about out-groups, but they did affect their perceptions of norms regarding the acceptability of openly engaging in prejudicial behavior. Furthermore, by affecting listeners' perceptions of social norms, their behavior was altered. That is, people exposed to the anti-prejudice program behaved in a less prejudiced way, even though they had not become less prejudiced.

A Trump Effect?

Overall, it is clear that people do take cues from peers or even media programs when deciding the appropriate way to behave when it comes to expressing their biases towards out-groups. Based on this literature and what we know about elite influence in general, it is reasonable to assume that political elites would also be able to send influential signals about the appropriate norms regarding expressions of prejudice. After all, the large body of scholarship on framing, priming, and agenda setting has demonstrated that elites can influence what people think and how they think about it when it comes to politics (Chong and Druckman, 2007; Schaffner and Sellers, 2009). At the same time, there is reason to think that elite influence may be limited in the particular realm related to expressions of prejudice. Indeed, Mendelberg (2001) shows that the public actually helps to enforce norms regarding expressions of prejudice by penalizing elites who use explicitly racist campaign appeals.

Yet, there have been several violations of the “norm of racial equality” in recent years, with the most notable example coming in the form of Donald Trump’s presidential campaign. Trump’s entry into the presidential race itself included a rant targeting Mexicans (quoted in the following section) that was widely criticized as offensive and beyond the pale of normal political discourse, even by many of his fellow Republicans. This speech was a sign of what was to come, as Trump frequently appeared to violate norms of civil discourse during the

campaign, particularly when it came to comments relating to various identity groups (Gross and Johnson, 2016). Yet, contrary to the expectations laid out by Mendelberg’s (2001) research, these remarks (and the subsequent criticism) did not derail his campaign. Notably, the fact that Trump did not suffer a significant penalty for his many prejudiced remarks is consistent with more recent experimental evidence that the public no longer penalizes elite expressions of explicit prejudice in the same way that it once did (Valentino, Neuner, and Vandebroek, 2018).

One culprit for these shifting patterns may be the rise of negative partisanship in contemporary American politics (Iyengar, Sood, and Lelkes, 2012). Partisans have developed increasingly negative emotions about the opposite party over the course of the past two decades, leading to a process of affective polarization. This affective polarization may be driven by the fact that the public’s partisan identities increasingly overlap with their social identities, such as race, ethnicity, and religion (Mason, 2016). As Mason and Wronski (2018) note, “the convergence of social identities along partisan lines makes in-party preference more powerful and out-party tolerance ever more difficult.” A potential consequence of such a process is to make the public particularly susceptible to influence from elite rhetoric (Tajfel and Turner, 1979).

There is some evidence that Trump’s election did alter people’s perceptions of norms regarding prejudice. Crandall, Miller, and White (2018) interviewed a convenience sample of individuals before and after the 2016 presidential election. They find that participants reported an increase in the perceived acceptability of expressing prejudice towards identity groups that Trump had targeted during the campaign even though the respondents did not themselves express higher levels of prejudice. Of course, it is one thing to find that people believe that the norms about expressions of racial prejudice have shifted, but what is especially important is to understand whether these shifting perceptions of norms have affected how people actually behave. Specifically, does Trump’s prejudiced rhetoric mean

that people feel more at liberty to express their own prejudices?

In the following section, I describe my experiments in which exposure to prejudicial remarks from Trump is randomized. I expect that seeing a prejudicial statement from Trump will lead people to say more negative (and offensive) things about identity groups because of the message those quotes send about the boundaries of acceptable discourse.

Design

Both experiments were conducted on nationally representative samples fielded by YouGov. The first experiment was fielded on the 2016 Cooperative Congressional Election Study (CCES) pre-election questionnaire and was administered to 1,186 non-Latino white adult respondents who were interviewed between September 28th and November 6th. The second experiment was fielded on 656 non-Latino non-Muslim white respondents who were interviewed between November 8th and December 9th, 2017, also as part of the CCES.

The primary aim of both experiments was to test whether respondents who were exposed to Trump’s offensive comments about minority groups would themselves express more negative sentiments about those groups. However, it was also important to disguise (as much as possible) the point of the experiment in order to minimize demand effects. As a result, exposure to the treatments was embedded in a question that was ostensibly seeking to understand how much attention respondents had paid to the presidential campaign.¹ Specifically, in the 2016 experiment the question asked respondents to identify which candidate made each of a series of statements during the presidential campaign. The preamble for this question read: “Now we will show you several statements made by the presidential candidates during the previous year. Please indicate which candidate you think made these statements.”

The control group saw just the following three statements:

¹Screen shots of the questions used in this experiment can be viewed in the Supplementary Information.

1. “Obama has no solutions. Obama has failed the country and its great citizens.”
2. “If I want to knock a story off the front page, I just change my hairstyle.”
3. “My two secrets to staying healthy: wash your hands all the time. And, if you can’t, use Purell or one of the sanitizers. And the other is hot peppers. I eat a lot of hot peppers. I for some reason started doing that in 1992, and I swear by it.”

Respondents saw one statement at a time and were asked to indicate whether Hillary Clinton or Donald Trump had made the remark, or if they were not sure. There were three treatment groups. Two treatment groups received one each of the following two statements, while the third treatment group received both additional quotes:

4. “Our great African-American President hasn’t exactly had a positive impact on the thugs who are so happily and openly destroying Baltimore.”
5. “When Mexico sends its people, they’re not sending their best. They’re sending people that have lots of problems. ... They’re bringing drugs. They’re bringing crime. They’re rapists. And some, I assume, are good people.”

Note that the order in which the quotes were presented was randomized for all respondents.

Following this question, on the next page, all respondents were asked “In a few words, please let us know what comes to mind when you think of the following groups?” Respondents were shown six groups and for each there was a small text box in which they could type out their responses. Respondents were asked to comment on Blacks, Mexicans, Whites, Politicians, the Middle Class, and Millennials (the order in which the groups appeared was randomized). While the target groups of interest for this experiment are Blacks and Mexicans (the targets of Trump’s offensive comments), the additional groups were included for two reasons. First, the additional groups helped to further disguise the purpose of this question vis-a-vis the experiment. Second, by including these additional groups, I can test for whether Trump’s offensive comments actually make respondents speak more negatively

about a broader array of groups than were targeted in Trump’s remarks. For the analysis that follows, I also analyze comments about Millennials for this purpose.

The 2017 experiment followed this same general format, but exposed respondents to one fewer control condition quotation. In particular, the quote that “Obama has no solutions...” was dropped from the experiment. More importantly, I used different quotations as treatments in this experiment:

3. “If you have people coming out of mosques with hatred and death in their eyes and on their minds, we’re going to have to do something.”
4. “You could put half of Trump’s supporters into what I call the ‘basket of deplorables’.
... The racist, sexist, homophobic, xenophobic, Islamophobic – you name it.”

The first quotation is another from Trump during the 2016 campaign, but this time targeting Muslims. Thus, including this quotation allows for a test of how people react when a different group is targeted. The second quote actually comes from Hillary Clinton. I include this quote to test whether calling out prejudice helps to mitigate or reduce the extent to which people are willing to express prejudice about other groups (Munger, 2017).

Respondents were shown seven groups after being exposed to the quotations and for each there was a small text box in which they could type out their responses. In the 2017 experiment, respondents were asked to comment on Muslims, Blacks, Whites, Politicians, the Middle Class, Republicans, and Democrats (the order in which the groups appeared was randomized). In both experiments, the prompt specifically asked respondents to provide just “a few words” and they were limited to 244 characters, which means many of the responses retrieved are quite short. In fact, the median response for comments about blacks in the 2016 experiment is just 17 characters in length.

Coding Respondent Comments

The open-ended responses represent the dependent variable of interest for this paper. However, the challenge is determining the best approach to coding them. Because the length of these comments is generally very brief, common automated text analysis tools struggle to extract meaningful information from them. Many respondents entered just one word in the boxes, such as “disadvantaged,” “entitled,” or “misunderstood.” Without incorporating the context of the question being answered, it is difficult to automate the categorization of some of these comments as negative or positive in nature. For example, the word “disadvantaged” has a negative tone, but its use here is not suggesting that the individual is saying something negative about Blacks. Similarly, the word “entitled” has a positive tone absent context, but in this use the respondent is clearly saying something negative about Blacks.

Thus, to code the comments from these survey responses, I relied on human coders recruited through Amazon’s Mechanical Turk (MTurk). Specifically, I recruited MTurk workers to code open ended comments about Blacks, Mexicans, and Millennials for the 2016 experiment and to code comments about Blacks, Whites, and Muslims for the 2017 experiment. Each worker was shown 10 comments, which were randomly assigned from the full list of comments. For the 2016 data, the median comment about Blacks was rated by 7 MTurk workers, and the median comment about Mexicans and Millennials was coded by 6 workers.² Each of the comments for the 2017 experiment was rated by a median of 9 MTurk workers. There was an attention check comment provided to each coder to make sure that the coders were actually reading the comments. Fewer than 2% of the MTurk coders failed the attention check; I exclude the ratings from this small number of MTurk coders who did not pass the attention check.

The MTurk workers were told what the question was that people were responding to

²After having MTurk workers code the comments about Blacks, I was able to increase the efficiency of the process by collapsing identical comments about Mexicans before placing them on MTurk.

and were asked to rate each comment on a 100-point scale ranging from “very negative” to “very positive.” Notably, the human coders appeared to be successful in using contextual information to discern the meaning of even brief text. Returning to the example above, the comment “entitled” received an average rating of 24 on the scale (clearly negative), while “disadvantaged” received a 43 (close to neutral). After rating each comment on the 100-point scale, MTurk workers were then asked to indicate whether the comment was hateful, offensive, or neither. This question was designed to help identify remarks that were especially negative in a way that MTurk workers would perceive as violating norms of polite discourse. For the analysis that follows, and consistent with Nithyanand, Schaffner, and Gill (2017), I combine the offensive and hateful categories together and refer to such remarks as offensive.

In the 2016 experiment, the average comment about Blacks received a rating of .54 while the average rating for a comment about Mexicans was .44 and for Millennials it was .60. On average, comments about Blacks were identified as offensive 29% of the time, compared to 20% for comments about Mexicans and 22% for comments about Millennials. In the 2016 experiment, the average negativity rating for blacks was .51, for Muslims it was .55, and for whites it was .49. Comments about Blacks were coded as offensive 26% of the time, compared to 32% for Muslims and 17% for Whites. For more information about the coding of comments, descriptive statistics of the codes, and examples of comments that were given the most negative ratings, see the Supplementary Information.

Accounting for Variance in Comment Coding

Since each comment was coded by multiple MTurk workers, the first dependent variable is the average rating given to each comment while the second dependent variable is the proportion of MTurkers who rated each comment as offensive or hateful. Of course, there is unequal variance in how each comment is rated. For example, in the 2016 experiment, 42% of comments about Blacks were coded as neither hateful nor offensive by all coders, while an

additional 6% were coded as offensive by all coders. But that means that for the remaining 52% of comments, at least one coder disagreed about whether the comment was offensive. Given the subjectivity of these assessments, it is not surprising to find disagreement on how comments are rated. However, it is important to account for this uncertainty when estimating the treatment effects. After all, a comment that was given a negativity rating of .5 by 6 coders indicates that the rating has a higher level of precision than one that was rated as a .25 by 3 coders and .75 by three other coders.

To account for this unequal variance across observations, I estimate my treatment effects using a bootstrapping approach. Specifically, I simulate 10,000 random draws for each sample of CCES respondents. For each dependent variable, I am able to specify not only the point estimate for the observation on that measure, but also the variance of the measure for each observation. Thus, for observations where the dependent variable is measured with less variance, the simulations will produce a smaller distribution of values around the point estimate across the 10,000 draws; but for observations with higher variance, this distribution will be larger (i.e. noisier). I then estimate the average treatment effect across the simulations, as well as a 95% confidence interval based on the distribution of treatment effects produced by the simulations. Again, the value of this approach is that it can incorporate the fact that the dependent variables are measured with varying levels of precision across observations.

Results

Figure 1 shows the treatment effects on negativity and offensiveness from the 2016 experiment. The figure shows the difference in the negativity of the comments offered by respondents receiving each comment relative to those that did not receive the comment. The horizontal lines are 95% bootstrapped confidence intervals around each treatment effect esti-

mate. The numerical values of the estimates plotted in the figure can be found in Supporting Information. First, looking at the left-hand plot, we see that neither of the quotes has a statistically significant effect on the negativity of comments offered about Blacks. The middle plot presents the treatment effects on the negativity of comments offered about Mexicans. Note that in this case there is a statistically significant treatment effect for receiving Trump's quote about Mexicans. Specifically, individuals who received the quote about Mexicans provided open ended comments about that group that were about 4 points more negative, on average, than those who did not receive the Trump quote. Once again, Trump's quote about thugs did not have a statistically significant effect on negativity towards Mexicans.

Finally, the top right-hand plot shows the tone of comments provided about Millennials. Interestingly, the patterns here are similar patterns to those in the previous plot. Specifically, being exposed to Trump's quote about Mexicans makes respondents more likely to offer negative comments about Millennials, while the quote about African Americans does not have a statistically significant impact the nature of these remarks.

The bottom row of plots in Figure 2 presents the treatment effects on the extent to which an individual's comment about Blacks, Mexicans, and Millennials was viewed as offensive or hateful. In this figure, we see consistent statistically significant treatment effects for the quotation about Mexicans on the offensiveness of comments offered about each group. Specifically, individuals exposed to the Trump quote about Mexicans were about 2 points more likely to write something offensive about Blacks, 5 points more likely to write something offensive about Mexicans, and four points more likely to write something offensive about Millennials. In each case, these treatment effects were statistically significant and substantively meaningful. For example, in the control condition, about 16% of the comments made about Mexicans were offensive; thus, a five percentage point increase in offensiveness means a nearly one-third increase over the baseline level of offensiveness.

By comparison, the effects of being exposed to Trump's quote about African Americans

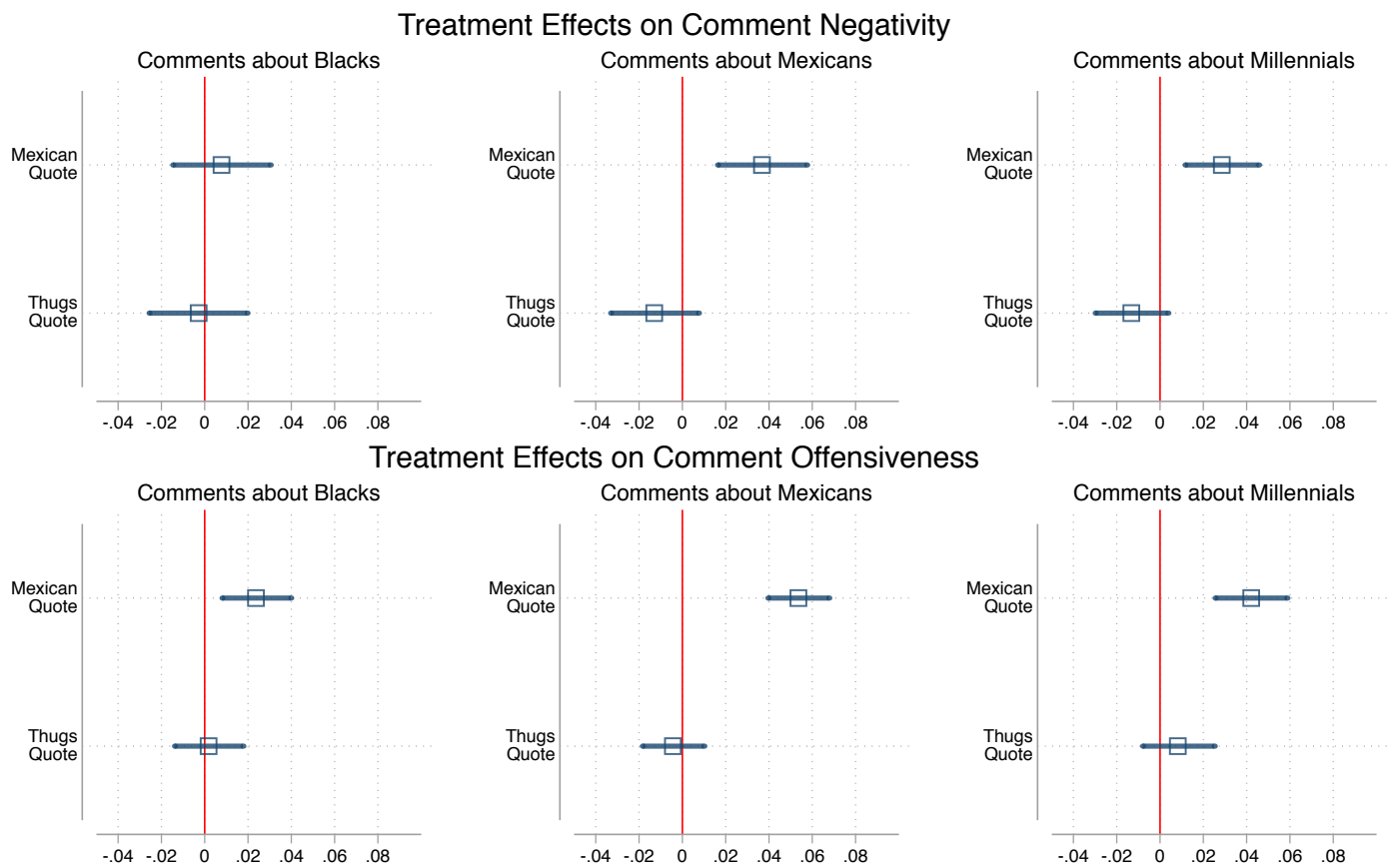
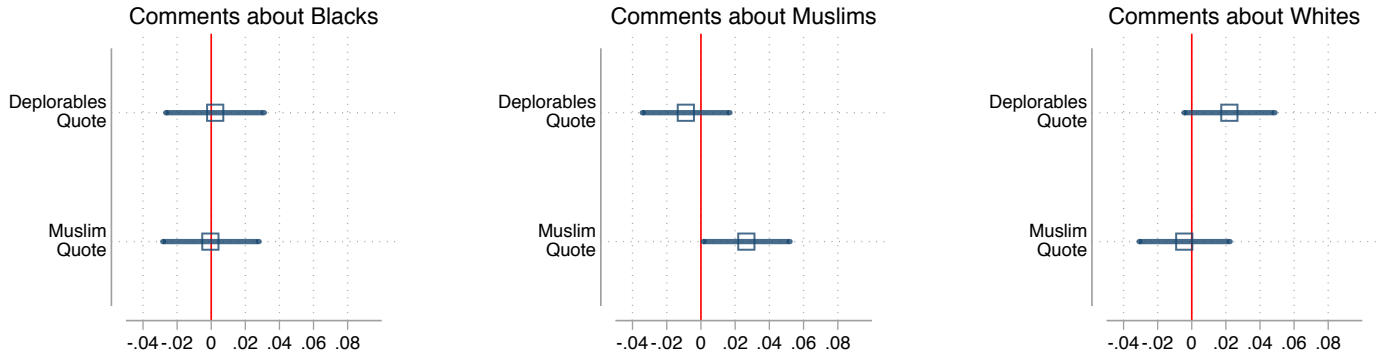


Figure 1: Treatment effects from experiment conducted in 2016. Treatment effects are the difference between ratings of comments from respondents in conditions receiving negative quotation about target group and those not receiving that quotation. Horizontal lines represent 95% confidence intervals.

Treatment Effects on Comment Negativity



Treatment Effects on Comment Offensiveness

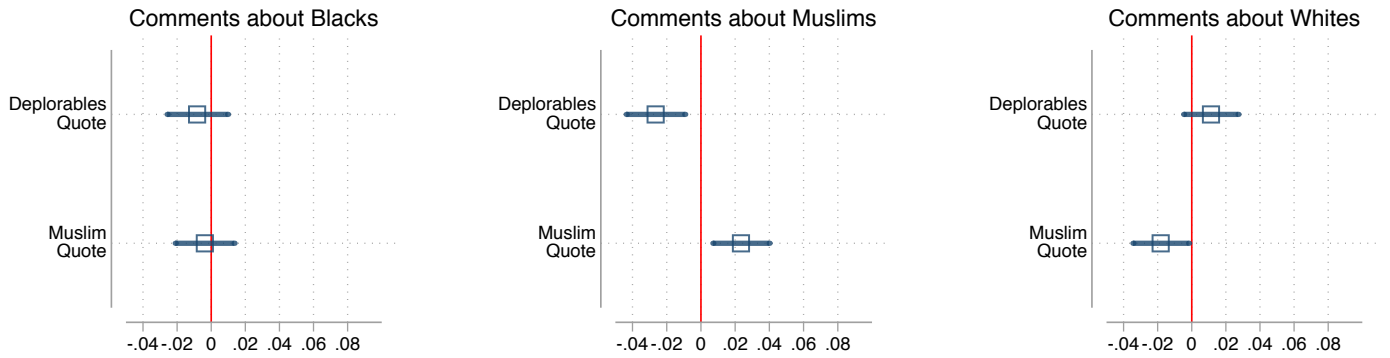


Figure 2: Treatment effects from experiment conducted in 2017. Treatment effects are the difference between ratings of comments from respondents in conditions receiving negative quotation about target group and those not receiving that quotation. Horizontal lines represent 95% confidence intervals.

resulted in small treatment effects that were not statistically significant for any of the three groups. Indeed, there are two patterns that are clear from the results. First, it is Trump's quote about Mexicans (rather than his quote about African Americans) that appears to influence how respondents talk about other groups in their own remarks. This pattern is likely due to the fact that the quote about Mexicans is especially offensive. Second, the effects of Trump's quote about Mexicans does not appear to be confined to the way in which respondents talked about Mexicans. Respondents who received this quote were also more likely to say negative things about Millennials. This suggests that norm-breaking rhetoric from elites might have consequences beyond how citizens talk about the particular group that was targeted by the elite.

Figure 2 shows the results from this second experiment. Note that for this experiment I analyze comments directed towards Blacks, Muslims, and Whites. I test for comments about Muslims since that group is the target of Trump's quote and I include whites since Clinton's quote is at least implicitly about white Americans. The top row of results show the effect of seeing each quote on the negativity of the comment an individual wrote about each group. The only statistically significant effect we see in these plots is for Trump's Muslim quote on the negativity of comments made about Muslims. Specifically, being exposed to Trump's quote increased the negativity of what the individual wrote about Muslims by 2.7 points.

The bottom row of Figure 2 shows the effects of being exposed to each quotation on the offensiveness of the comments written about each group. As with negativity, we see no statistically significant effect for comments about Blacks. However, we do see clear movement when it comes to the offensiveness of comments made about Muslims. Individuals who were exposed to Trump's negative quote about Muslims were 2.3 points more likely to say something offensive about Muslims. Notably, however, exposure to Clinton's deplorable quotation reduced the propensity of individuals to say something offensive about Muslims by 2.6 points. This provides some evidence that rhetoric condemning prejudice may be effective

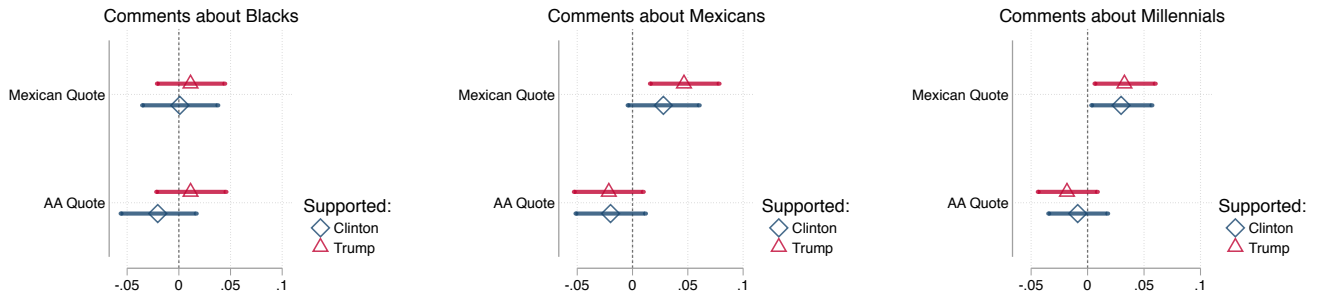
for reducing offensive speech about minority groups. Also of note, the bottom-right plot shows that Trump’s quote about Muslims actually led to a slight decrease in the propensity of respondents to say something negative about Whites.

Figure 3 displays the treatment effects for the 2016 experiment for respondents who voted for Trump versus Clinton separately.³ While there are no statistically significant differences in treatment effects for Trump voters relative to Clinton voters when it comes to the negativity of comments made about the groups, there are some significant and notable differences when it comes to offensiveness. In most cases Trump voters were more likely than Clinton voters to react to the Trump quotations by making more offensive comments themselves. However, there are two exceptions to this pattern. First, Trump voters were actually less likely to say something offensive about Mexicans when they were exposed to the “thugs” quote about African Americans, whereas that quote had no effect on Clinton supporters. Second, exposure to Trump’s quote about Mexicans increased the offensiveness of statements made about Millennials among both Clinton and Trump voters by a similar magnitude.

In the 2017 experiment, there were no clear or consistent patterns of the quotations affecting Trump supporters differently from Clinton supporters. The Supplementary Information has more detail on this analysis. Thus, overall, while exposure to Trump’s offensive quotations sometimes affected Trump supporters more than Clinton supporters, this was not always the case, indicating that these statements have the potential to influence people beyond Trump’s base of supporters.

³I coded an individual as voting for Trump if they said that they intended to vote for him or said that they were not going to vote but that they preferred him to the other candidates. I coded Clinton voters in the same way.

Treatment Effects on Comment Negativity



Treatment Effects on Comment Offensiveness

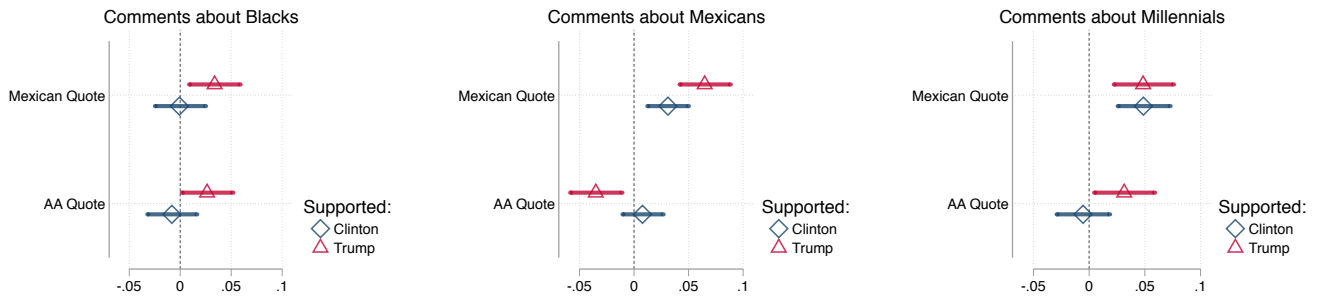


Figure 3: Treatment effects on negativity and offensiveness of open-ended comments by 2016 presidential vote/preference. Treatment effects are the difference between ratings of comments from respondents in conditions receiving negative quotation about target group and those in the control condition. Horizontal line represents 95% confidence intervals.

Is it Mimicry?

While I theorize that the findings presented above result from the altering of perceived norms about expressions of prejudice, there is an alternative mechanism through which we might also observe these results. Specifically, the quotations might simply make particular words more accessible in the minds of individuals when they are asked to evaluate the same (or similar) groups. In this section, I test for this alternative mechanism by examining the incidence with which respondents used words that they were exposed to in Trump’s quotes. The results from this analysis are presented in Table 1. The first row indicates the percentage of respondents who used a word to describe either blacks, Mexicans, or Millennials that were also words that Trump used in his quote about Mexicans (e.g. “drugs,” “crime,”

Terms	Not exposed to quote	Exposed to quote	Difference
rap*/drug*/crim*	4.4%	5.6%	1.2%
	(0.9)	(0.9)	(1.3)
thug*	0.3%	1.5%	1.2%
	(0.2)	(0.5)	(0.6)

Table 1: Respondents’ Usage of Words from Trump Quotations, 2016 Experiment. Entries are percentage of respondents using a word with the indicated stems. Columns identify whether respondents were exposed to a Trump quotation that included words with those same stems. Standard errors in parentheses.

and “rapists”). 4.4% of respondents who were not exposed to Trump’s quote about Mexicans used words that Trump used in his quotations about Mexicans, whereas 5.6% of those who received Trump’s quote used these words in their own remarks. Thus, those receiving the Trump quote were 1.2 points more likely to use words that appear in that quote; however, this difference is relatively small and not statistically significant.

The second row of results in Table 1 compares the incidence of the use of the stem “thug” among respondents who did and did not receive Trump’s quote referring to African Americans as “thugs.” Once again, we see that those receiving Trump’s quote used the term 1.2 points more frequently than those who did not. However, in this case the difference is statistically significant ($p = .04$).

Overall, then, the results in Table 1 suggest that there may be some minimal degree of mimicry taking place among respondents exposed to Trump’s quotes. However, the magnitude of these effects are small, suggesting that mimicry alone cannot explain the effects we find for the influence of Trump’s rhetoric on statements made by the subjects in this study.

Table 2 presents a similar analysis for the 2017 experiment. The differences in word usage for those who did and did not receive the quotations is very small and not statistically significant. Overall, then, the results in Tables 1 and 2 suggest that there may be some minimal degree of mimicry taking place among respondents exposed to Trump’s quotes. However, the magnitude of these effects are small, suggesting that mimicry alone cannot

Terms	Not exposed to quote	Exposed to quote	Difference
hate/hatred/death	3.4%	3.3%	0.1%
	(1.0)	(1.0)	(1.4)
deplor*/rac*/sex*/homophobi*/ islamaphobi*/xenophobi*	10.3%	10.7%	0.3%
	(1.7)	(1.7)	(2.4)

Table 2: Respondents’ Usage of Words from Trump/Clinton Quotations, 2017 Experiment. Entries are percentage of respondents using a word with the indicated stems. Columns identify whether respondents were exposed to a quotation that included words with those same stems. Standard errors in parentheses.

explain the effects we find for the influence of Trump’s rhetoric on statements made by the subjects in this study.

Group Contact and Signals About Norms

As noted above, people may have different understandings of the norms regarding expressions of prejudice and, likewise, people may be more or less confident in what those norms are. One factor that may affect an individual’s perceptions of norms (and their confidence in those perceptions) is the extent to which they interact with members of minority groups. Contact theory predicts that individuals who interact with members of other groups will hold less prejudice toward those groups as a result of those interpersonal interactions (Allport, 1954). Studies that have tested this prediction have generally found that people do, in fact, express less prejudice towards groups when they experience more contact with members of those groups (Pettigrew and Tropp, 2006). It is natural to expect that contact with members of an out-group might not only reduce the extent to which one holds prejudice towards that group, but it might also make people less likely to express whatever prejudices they do have.

Thus, I expect that exposure to Trump’s prejudiced comments will be less influential on what white respondents say about Blacks and Mexicans when those respondents talk about politics with members of either group. To test this expectation, I draw on a question from the CCES module that asked individuals to indicate how many people they talked with about

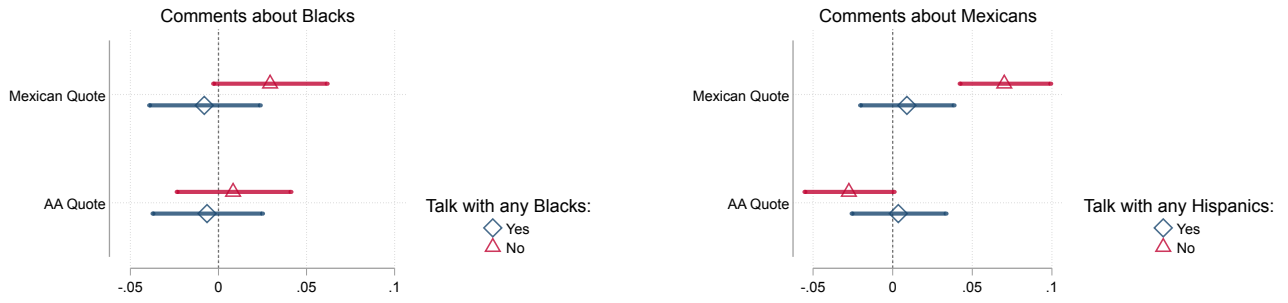
politics or the election from a variety of different identity groups.⁴ I focus on the items related to conversations with Blacks and with Hispanics (respondents were not asked whether they had discussions with Millennials). In the latter case, of course, contact with Hispanics is not a perfect corollary to the question about Mexicans, but it seems reasonable to assume that contact with Hispanics would reduce expressions of prejudice towards Mexicans. Notably, 46% of whites in the sample reported that they had not talked about politics or the election with any blacks and 52% said that they had not had such discussions with any Latinos. These items were only asked in the 2016 survey, so I limit this part of the analysis to that experiment.

Figure 4 plots the treatment effects of being exposed to each Trump quotation on the tone and offensiveness of the comments offered about Blacks and Mexicans. Overall, the patterns in the figure indicate that Trump’s quote about Mexicans was particularly influential in affecting what whites said about Blacks and Mexicans when those individuals reported that they had not discussed politics with any blacks or Latinos. For example, whites who had no discussions with blacks reacted to Trump’s quote about Mexicans by saying things about blacks that were about 3 points more negative and about 4 points more offensive. Exposure to the Trump quotation had a small and statistically indistinguishable effect on what whites said about blacks when they had talked with blacks about the election.

The right-hand plots in Figure 4 indicate that exposure to Trump’s quotation had even larger effects on what people wrote about Mexicans when they had not discussed politics with any Hispanics. Whites who had no contact with Hispanics made comments about Mexicans that were about 7 points more negative and 8 points more offensive when they were exposed to the Trump quote about Mexicans. These effects were much larger than those recorded

⁴The actual question was, “Thinking about all of the people with whom you have discussed politics or the upcoming election over the past month, about how many fall into the following categories? This includes any discussion of the election, specific candidates, political issues, or the performance of local, state or national government. Political discussions might have occurred in person, on the phone, by email, over social media, or in online discussion forums.”

Treatment Effects on Comment Negativity



Treatment Effects on Comment Offensiveness

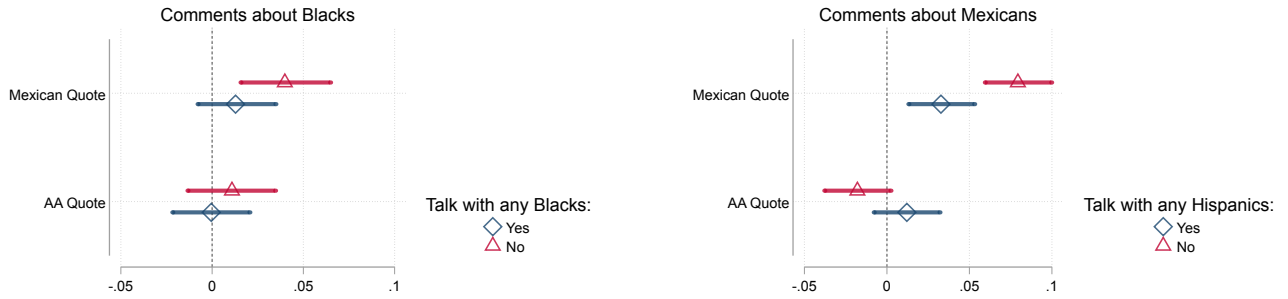


Figure 4: Treatment effects on negativity and offensiveness of open-ended comments by contact with Blacks or Hispanics. Treatment effects are the difference between ratings of comments from respondents in conditions receiving negative quotation about target group and those in the control condition. Horizontal line represents 95% confidence intervals.

for individuals who did report discussing politics with at least one Hispanic person.

As with the results presented above, Trump’s quote about African Americans was less influential on how people talked about Blacks and Mexicans. Exposure to that quotation does not produce a statistically significant increase in negativity of offensiveness toward either groups, regardless of whether the individual had been in contact with members of that group.

Conclusion

This paper presented an experimental test of the “Trump Effect” – the notion that Trump’s offensive and prejudicial rhetoric on the campaign trail might cause individuals to express more prejudice toward out-groups. I find support for this expectation; specifically, individuals who were exposed to Trump’s quote about Mexicans were significantly more likely to make negative and offensive remarks not only about Mexicans, but also about other identity groups such as Blacks and Millennials. The fact that I uncover significant effects for exposure to the quotation is especially noteworthy given the fact that many respondents had undoubtedly already heard Trump’s quote before and may have already adjusted their expressions of prejudice accordingly. The likelihood that many had been “pre-treated” means that the effects uncovered here may represent a conservative estimate of the actual impact of Trump’s rhetoric (Gaines, Kuklinski, and Quirk, 2006).

While it is impossible to know the true mechanism by which Trump’s prejudiced remarks influenced how people talked about out-groups, the evidence presented here is consistent with the notion that individuals took the offensive quotations as a cue about the norms regarding expressions of prejudice. First, very little mimicry was detected, indicating that people were not simply re-using the words that they had just heard. Second, exposure to the offensive quotation had the strongest impact among individuals who reported having

no political discussions with the targeted out-groups. It is these individuals who are likely to hold higher levels of prejudice and who are likely to have less clarity on what types of comments are beyond the realm of polite discourse. Finally, the notion that Trump's quotation sends a signal about norms is consistent with research by Crandall, Miller, and White (2018) showing that people's perceptions of norms regarding expressions of prejudice were affected by Trump's election. Indeed, the authors find that "Supporters of Trump and Clinton alike saw increased approval for expression of prejudices that characterized the Trump campaign" (p. 4).

Several studies have already demonstrated that racial resentment played a larger role in predicting vote choices in 2016 than it had in previous elections (Schaffner, MacWilliams, and Nteta, 2018; Sides, Tesler, and Vavreck, 2018). Indeed, Trump's success in winning the presidency can be partly attributed to the fact that citizens no longer appear to be motivated to punish candidates who make explicitly prejudiced statements (Valentino, Neuner, and Vandebroek, 2018) and that racial/ethnic identity groups are increasingly conflated with partisan identity (Mason, 2016). The result is that Republican candidates increasingly have an incentive to use prejudiced rhetoric to appeal to their supporters, and using such rhetoric no longer comes with the same electoral penalty as it once did. This study has demonstrated how such a strategy could have profound implications for the nature of mass discourse. Indeed, if politicians increasingly feel at liberty to use explicitly prejudiced rhetoric during their campaigns, then the mass public is likely to take cues from such behavior, leading them to express more prejudice themselves. The result would almost certainly be increasingly heightened inter-group tensions which pose a threat to political and social stability in the United States. Importantly, however, the evidence from the second experiment suggests that the sanctioning of prejudice by Clinton was able to mitigate the effects of Trump's prejudiced statements. This is consistent with other research that also finds that sanctioning of expressions of prejudice can be effective (Munger, 2017). Thus, to avoid the political and

social consequences that may result from increasing expressions of prejudice among the mass public, it is crucial that members of the political and media elite engage in strong and clear sanctioning behavior whenever politicians do use prejudiced rhetoric. The hope is that such messages will help to counteract the chilling effects presented in this study.

References

- Allport, Gordon W. 1954. *The nature of prejudice*. Addison-Wesley.
- Blanchard, Fletcher A, Christian S Crandall, John C Brigham, and Leigh Ann Vaughn. 1994. "Condemning and condoning racism: A social context approach to interracial settings." *Journal of Applied Psychology* 79(6): 993.
- Blanchard, Fletcher A, Teri Lilly, and Leigh Ann Vaughn. 1991. "Reducing the expression of racial prejudice." *Psychological Science* 2(2): 101–105.
- Chong, Dennis, and James N Druckman. 2007. "Framing theory." *Annu. Rev. Polit. Sci.* 10: 103–126.
- Cialdini, Robert B, and Melanie R Trost. 1998. "Social influence: Social norms, conformity and compliance."
- Costello, Maureen B. 2016. "The Trump effect: The impact of the presidential campaign on our nation's schools." *Montgomery, AL: Southern Poverty Law Center* .
- Crandall, Christian S, and Amy Eshleman. 2003. "A justification-suppression model of the expression and experience of prejudice." *Psychological bulletin* 129(3): 414.
- Crandall, Christian S, Jason M Miller, and Mark H White. 2018. "Changing Norms Following the 2016 US Presidential Election: The Trump Effect on Prejudice." *Social Psychological and Personality Science* p. 1948550617750735.
- Crosby, Faye, Stephanie Bromley, and Leonard Saxe. 1980. "Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review." *Psychological Bulletin* 87(3): 546.

- Dovidio, John F, and Samuel L Gaertner. 1986. *Prejudice, discrimination, and racism*. Academic Press.
- Gaines, Brian J, James H Kuklinski, and Paul J Quirk. 2006. "The logic of the survey experiment reexamined." *Political Analysis* 15(1): 1–20.
- Gervais, Bryan T. 2014. "Following the news? Reception of uncivil partisan media and the use of incivility in political expression." *Political Communication* 31(4): 564–583.
- Gervais, Bryan T. 2015. "Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment." *Journal of Information Technology & Politics* 12(2): 167–185.
- Gross, Justin H, and Kaylee T Johnson. 2016. "Twitter taunts and tirades: Negative campaigning in the age of Trump." *PS: Political Science & Politics* 49(4): 748–754.
- Hurwitz, Jon, and Mark Peffley. 1992. "Traditional versus social values as antecedents of racial stereotyping and policy conservatism." *Political Behavior* 14(4): 395–421.
- Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, Not Ideology: A Social Identity Perspective on Polarization." *Public opinion quarterly* 76(3): 405–431.
- Kuklinski, James H, Michael D Cobb, and Martin Gilens. 1997. "Racial attitudes and the "New South"." *The Journal of Politics* 59(2): 323–349.
- Kwon, K Hazel, and Anatoliy Gruzd. 2017. Is Aggression Contagious Online? A Case of Swearing on Donald Trump's Campaign Videos on YouTube. In *Proceedings of the 50th Hawaii International Conference on System Sciences*.
- Mason, Lilliana. 2016. "A cross-cutting calm: How social sorting drives affective polarization." *Public Opinion Quarterly* 80(S1): 351–377.

- Mason, Lilliana, and Julie Wronski. 2018. "One Tribe to Bind Them All: How Our Social Group Attachments Strengthen Partisanship." *Political Psychology* 39(S1): 257–277.
- Mendelberg, Tali. 2001. *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press.
- Munger, Kevin. 2017. "Tweetment effects on the tweeted: Experimentally reducing racist harassment." *Political Behavior* 39(3): 629–649.
- Nithyanand, Rishab, Brian Schaffner, and Phillipa Gill. 2017. "Online Political Discourse in the Trump Era." *arXiv preprint arXiv:1711.05303* .
- Paluck, Elizabeth Levy. 2009. "Reducing intergroup prejudice and conflict using the media: a field experiment in Rwanda." *Journal of personality and social psychology* 96(3): 574.
- Paluck, Elizabeth Levy, and Donald P Green. 2009. "Prejudice reduction: What works? A review and assessment of research and practice." *Annual review of psychology* 60: 339–367.
- Pettigrew, Thomas F, and Linda R Tropp. 2006. "A meta-analytic test of intergroup contact theory." *Journal of personality and social psychology* 90(5): 751.
- Rokeach, Milton, and Sandra J Ball-Rokeach. 1989. "Stability and change in American value priorities, 1968–1981." *American Psychologist* 44(5): 775.
- Schaffner, Brian F, and Patrick J Sellers. 2009. *Winning with words: the origins and impact of political framing*. Routledge.
- Schaffner, Brian F, Matthew MacWilliams, and Tatishe Nteta. 2018. "Understanding white polarization in the 2016 vote for president: The sobering role of racism and sexism." *Political Science Quarterly* 133(1): 9–34.

- Sides, John, Michael Tesler, and Lynn Vavreck. 2018. "Identity Crisis: The 2016 Election and the Battle for the Meaning of America."
- Tajfel, Henri, and John C Turner. 1979. "An integrative theory of intergroup conflict." *The social psychology of intergroup relations* 33(47): 74.
- Valentino, Nicholas A, Fabian G Neuner, and L Matthew Vandenbroek. 2018. "The Changing Norms of Racial Political Rhetoric and the End of Racial Priming." *The Journal of Politics* 80(3): 000–000.
- Zaller, John. 1992. *The nature and origins of mass opinion*. Cambridge university press.
- Zitek, Emily M, and Michelle R Hebl. 2007. "The role of social norm clarity in the influenced expression of prejudice over time." *Journal of Experimental Social Psychology* 43(6): 867–876.