Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media

Christopher Bail,[1*] Lisa Argyle,[2] Taylor Brown,[1] John Bumpus,[1]
Haohan Chen,[3] M.B. Fallin Hunzaker,[4] Jaemin Lee, [1] Marcus Mann,[1]
Friedolin Merhout,[1] Alexander Volfovsky[5]

[1]Department of Sociology, Duke University,
254 Soc.-Psych Bldg, Durham, NC 27708, USA
[2]Department of Politics, Princeton University,
130 Corwin Hall, Princeton, NJ 08544, USA
[3]Department of Political Science, Duke University,
140 Science Drive, Durham, NC 27708, USA
[4]Department of Sociology, New York University,
295 Lafayette Street, 4th Floor, New York, NY 10012, USA
[5]Department of Statistical Science, Duke University,
P.O. Box 90251, Durham, NC 27708, USA

[*]To whom correspondence should be addressed: christopher.bail@duke.edu.

There is mounting concern that social media sites contribute to political polarization by creating "echo chambers" that insulate people from opposing views about current events. We surveyed a large sample of Democrats and Republicans who visit Twitter at least three times each week about a range of social policy issues. One week later, we randomly assigned respondents to a treatment condition in which they were offered financial incentives to follow a

Twitter bot for one month that exposed them to messages produced by elected officials, organizations, and other opinion leaders with opposing political ideologies. Respondents were re-surveyed at the end of the month to measure the effect of this treatment, and at regular intervals throughout the study period to monitor treatment compliance. We find that Republicans who followed a liberal Twitter bot became substantially more conservative post-treatment, and Democrats who followed a conservative Twitter bot became slightly more liberal post-treatment. These findings have important implications for the interdisciplinary literature on political polarization as well as the emerging field of computational social science.

Political polarization in the United States has become a central focus of social scientists in recent decades (*1–7*). Americans remain deeply divided on controversial issues such as inequality, race, and immigration. According to the 2016 National Election Study, 59.3% of Clinton voters believe federal aid to the poor should be increased compared to only 20.2% of Trump voters. 77.7% of Clinton voters express favorable attitudes towards the Black Lives Matter movement, whereas 31.2% of Trump voters do the same. 68.9% of Trump voters believe immigration to the United States should be decreased, compared to 21.9% of Clinton voters. Longstanding divides about these and many other issues have far-reaching consequences for the design and implementation of social policies as well as the effective function of democracy more broadly (*8–12*).

America's deep partisan divides are often attributed to "echo chambers," or patterns of information sharing that reinforce pre-existing political beliefs by limiting exposure to heterogeneous ideas and perspectives (*13–17*). Concern about selective exposure to information and political polarization has increased in the age of social media (*13, 18–20*). The vast majority of Americans now visit a social media site at least once each day, and a rapidly growing number of them list social media as their primary source of news (*21*). Despite initial optimism that social media might enable people to consume more heterogeneous sources of information about current events, there is growing concern that such forums exacerbate political polarization because of social network homophily, or the well-documented tendency of people to form social network ties to those who are similar to themselves (*22, 23*). The endogenous relationship between social network formation and political attitudes also creates formidable challenges for the study of social media echo chambers and political polarization, since it is notoriously difficult to establish whether information sharing networks shape political opinions, or vice versa (*24–26*).

Here, we report the results of a large field experiment designed to examine whether disrupting selective exposure to partisan information on social media sites shapes political attitudes. Our research is governed by three pre-registered hypotheses. Our first hypothesis is that disrupting selective exposure to partisan information will decrease political polarization because of inter-group contact effects. A vast literature indicates contact between opposing groups can minimize perceived differences or stereotypes that develop in the absence of positive interactions between them (*27*). Studies also indicate inter-group contact increases the likelihood of deliberation and political compromise (*28, 29*). Yet recent large-scale field-experiments reveal evidence both for and against the hypothesis that inter-group contact improves inter-group attitudes (*30, 31*)— and to our knowledge, no studies have yet examined how this dynamic unfolds on social media.

Our second hypothesis builds upon a more recent wave of studies that indicate exposure to those with opposing political views may create backfire effects that exacerbate political polarization (*32–35*). This literature—which now spans several academic disciplines—indicates people who are exposed to messages that conflict with their own attitudes are prone to counter-argue them using motivated reasoning, which accentuates perceived differences between groups and increases their commitment to pre-existing beliefs (*32–35*). Many studies in this literature observe backfire effects via survey experiments where respondents are exposed to information that corrects factual inaccuracies—such as the notion that Iraq possessed weapons of mass destruction before the second Gulf War—though these findings have failed to replicate in two recent studies. (*36, 37*). Our study is not designed to evaluate attempts to correct factual inaccuracies, however, but the broader impact of prolonged exposure to counter-attitudinal messages on social media.

Our third pre-registered hypothesis is that backfire effects will be more likely to occur among conservatives than liberals. This hypothesis builds upon recent studies that indicate conservatives hold values that prioritize certainty and tradition whereas liberals value change and diversity (*38, 39*). It also builds upon studies that observe asymmetric polarization in roll-call voting wherein liberal politicians have gradually shifted towards the center across recent decades whereas conservatives have remained steadfast on the right (*40*). Once again, we are not aware of any other studies that have examined such dynamics among the broader public on social media.

Research Design

Figure 1 provides an overview of our research design. We hired a professional survey firm to recruit self-identified Republicans and Democrats from a large web-based panel who visit Twitter at least three times each week to complete a 10 minute survey in mid-October 2017. Our outcome of interest is change in political ideology, measured with a ten-item survey instrument in which respondents were asked to agree or disagree with a range of statements about policy issues on a seven-point scale ($\alpha = .91$) (*41*). Our survey also collected information about other political attitudes, use of social media and conventional media sources, and a range of demographic indicators that we describe in our Supplementary Materials. Finally, all respondents were asked to report their Twitter ID, which we used to mine additional information about their online behavior, including the partisan background of the accounts they follow on Twitter.

We ran separate field experiments for Democratic and Republican respondents, and within each group we employed a block randomization design that further stratified respondents according to two variables that have been linked to political polarization: a) level of attachment to political party, and, b) level of interest in current events. We also randomized assignment according to respondents' frequency of twitter usage, which we reasoned would influence the amount of exposure to the intervention we describe in the following paragraph and thereby the overall likelihood of opinion change.

We received 1,652 responses to our pre-treatment survey (901 Democrats and 751 Republicans). One week later, we randomly assigned respondents to a treatment condition, thus employing an "ostensibly unrelated" survey design (*31*). At this time, respondents in the treatment condition were offered $11 to follow a Twitter bot, or automated Twitter account, that they were told would retweet twenty-four messages each day for one month. Respondents were not informed of the content of the messages the bots would retweet. As Figure 2 illustrates, we created a liberal Twitter bot and a conservative Twitter bot for each of our experiments that retweeted messages that were randomly drawn from a sample of 4,176 political Twitter accounts (e.g. elected officials, opinion leaders, media organizations, and non-profit groups) identified via a network-sampling technique that we describe in further detail within our Supplementary Materials (*42*).

To monitor treatment compliance, respondents were offered additional financial incentives (up to $18) to complete weekly surveys that asked them to answer questions about the content of the tweets produced by the Twitter bots and identify a picture of an animal that was tweeted twice a day by the bot but deleted immediately prior to the weekly survey. At the conclusion of the study period, respondents were asked to complete a final survey with the same questions from the initial (pre-treatment) survey. Of those invited to follow a Twitter bot, 64.8 percent

Initial Survey

Respondents were offered $11 to provide their Twitter ID and complete a 10-minute survey about their political attitudes, social media use, and media consumption habits (demographics provided by survey firm)

Randomization

One week later, respondents were assigned to treatment and control conditions within strata created using pre-treatment covariates that describe attachment to party, frequency of Twitter use, and overall interest in current events.

Weekly Surveys

Respondents in treatment conditions informed they are eligible to receive up to $6 each week during the study period for correctly answering questions about the content of messages retweeted by Twitter Bots .

Post-Survey

Respondents were offered $12 to repeat the pre-treatment survey one month after initial survey

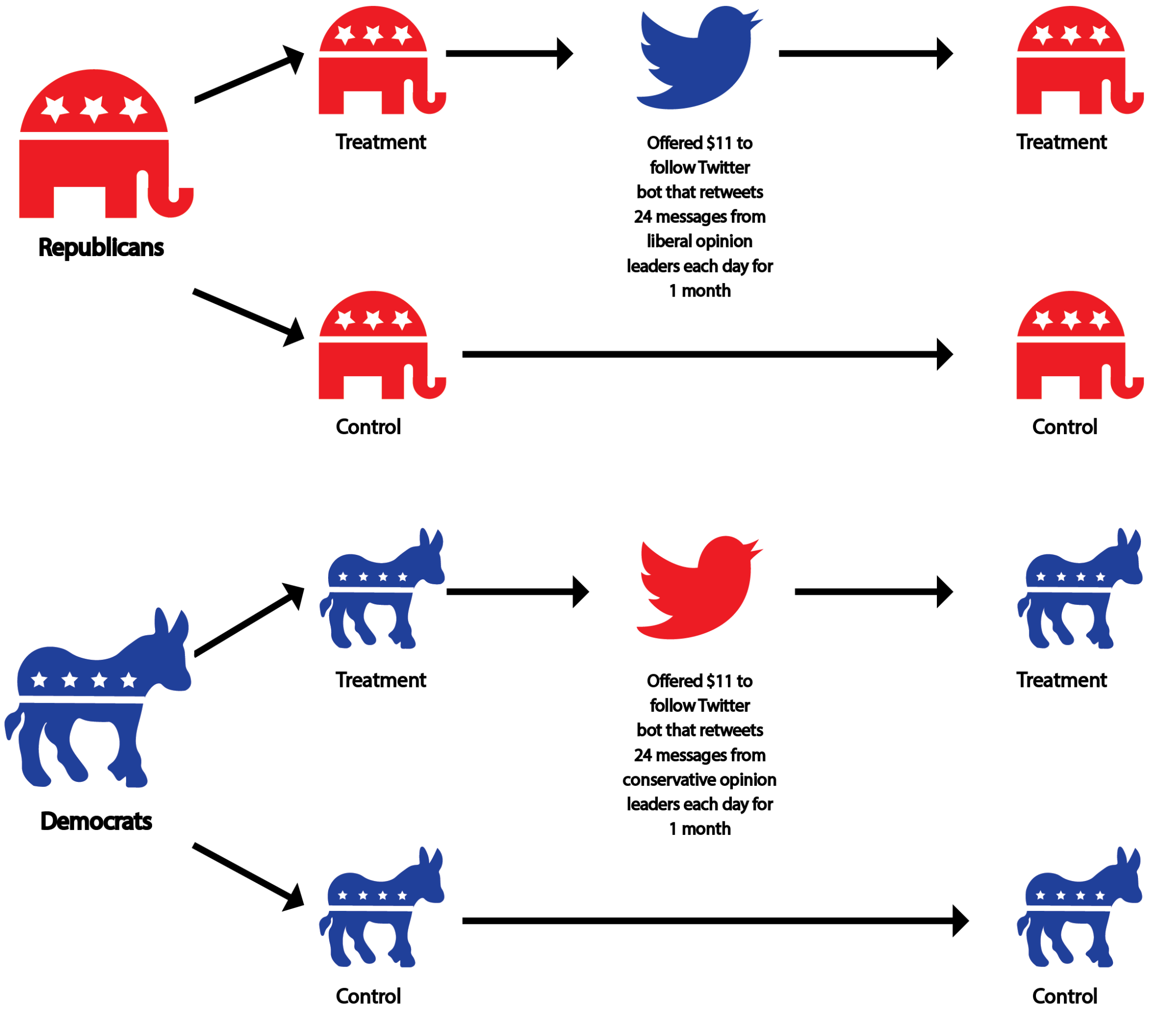
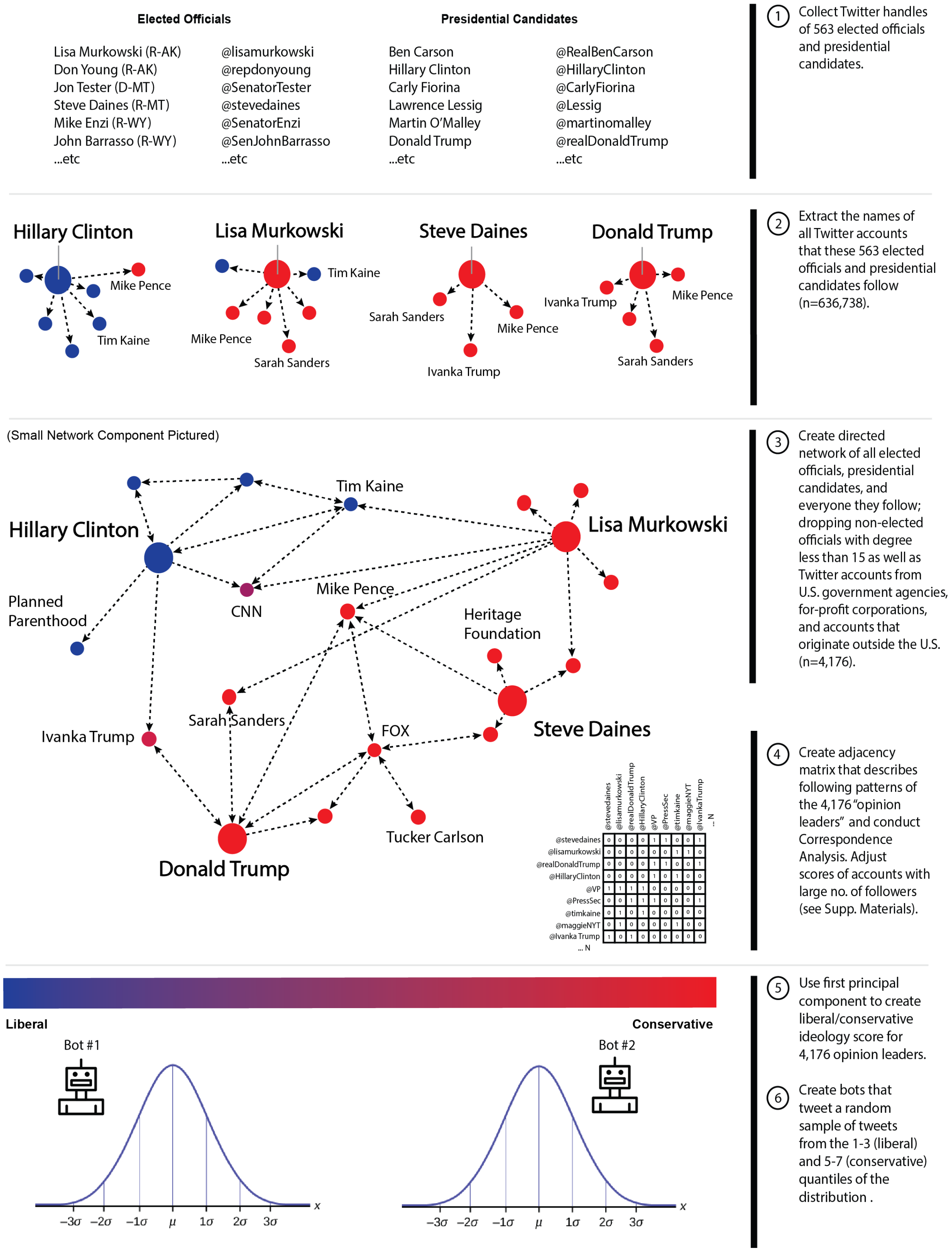

Republicans
Treatment
Offered $11 to follow Twitter bot that retweets 24 messages from liberal opinion leaders each day for 1 month
Treatment
Control
Control
Democrats
Treatment
Offered $11 to follow Twitter bot that retweets 24 messages from conservative opinion leaders each day for 1 month
Treatment
Control
Control

Figure 1: Overview of Research Design



Elected Officials		Presidential Candidates	
Lisa Murkowski (R-AK)	@lisamurkowski	Ben Carson	@RealBenCarson
Don Young (R-AK)	@repdonyoung	Hillary Clinton	@HillaryClinton
Jon Tester (D-MT)	@SenatorTester	Carly Fiorina	@CarlyFiorina
Steve Daines (R-MT)	@stevedaines	Lawrence Lessig	@Lessig
Mike Enzi (R-WY)	@SenatorEnzi	Martin O'Malley	@martinomalley
John Barrasso (R-WY)	@SenJohnBarrasso	Donald Trump	@realDonaldTrump
...etc	...etc	...etc	...etc

1 Collect Twitter handles of 563 elected officials and presidential candidates.

2 Extract the names of all Twitter accounts that these 563 elected officials and presidential candidates follow (n=636,738).

3 Create directed network of all elected officials, presidential candidates, and everyone they follow; dropping non-elected officials with degree less than 15 as well as Twitter accounts from U.S. government agencies, for-profit corporations, and accounts that originate outside the U.S. (n=4,176).

4 Create adjacency matrix that describes following patterns of the 4,176 "opinion leaders" and conduct Correspondence Analysis. Adjust scores of accounts with large no. of followers (see Supp. Materials).

5 Use first principal component to create liberal/conservative ideology score for 4,176 opinion leaders.

6 Create bots that tweet a random sample of tweets from the 1-3 (liberal) and 5-7 (conservative) quantiles of the distribution .

	@stevedaines	@lisamurkowski	@realDonaldTrump	@HillaryClinton	@VP	@PressSec	@timkaine	@maggieNYT	@IvankaTrump	... N
@stevedaines	0	0	0	0	1	1	0	0	1	
@lisamurkowski	0	0	0	0	0	0	1	1	0	
@realDonaldTrump	0	0	0	0	1	1	0	0	1	
@HillaryClinton	0	0	0	0	1	0	1	0	0	
@VP	1	1	1	1	0	0	0	0	0	
@PressSec	0	0	1	1	1	0	0	0	1	
@timkaine	0	1	0	1	0	0	0	0	0	
@maggieNYT	0	1	0	0	0	0	1	0	0	
@Ivanka Trump	1	0	1	0	0	0	0	0	0	
... N										

Figure 2: Overview of Research Design

of Democrats and 57.6 percent of Republicans accepted our invitation. 40.14 percent of respondents across both samples were able to answer all substantive questions about the content of messages retweeted each week and 32.48 percent were able to identify the animal picture re-tweeted each day.

Results

Figure 3 reports both the effect of being assigned to the treatment condition, or the Intent to Treat effect (ITT), as well as the Complier Average Causal Effects (CACE), which account for the differential rates of compliance among respondents we observed. These estimates were produced via multivariate models that predict respondents' post-treatment scores on the liberal/conservative scale described above that control for pre-treatment scores on this scale as well as twelve other covariates described in our Supplementary Materials. Negative scores indicate respondents became more liberal in response to treatment and positive scores indicate they became more conservative. Circles describe unstandardized point estimates and the horizontal lines in Figure 3 describe 90 and 95 percent confidence intervals. We measured compliance with treatment in three ways. "Minimally Compliant Respondents" describes those who followed our bot within five days after the invitation was made. "Partially Compliant Respondents" are those who were able to answer at least one—but not all—questions about the content of a bot's tweets administered each week during the survey period. "Fully Compliant Respondents" are those who successfully answered all of these questions.

Though treated Democrats exhibited slightly more liberal attitudes post-treatment (that increase in size with level of compliance), none of these effects were statistically significant. Treated Republicans, by contrast, exhibited substantially more conservative views post-treatment. These effects also increase with level of compliance, but they are highly significant at the $p <.001$ level. Our most conservative estimate is that treated Republicans increased .12 points

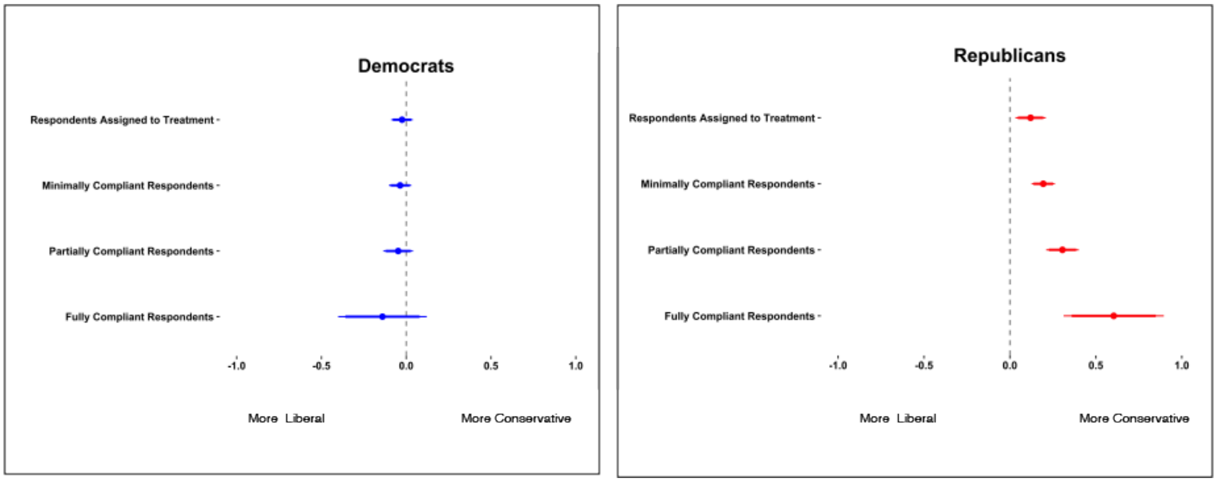


Figure 3: Effect of following Twitter bots that re-tweet messages by elected officials, organizations, and opinion leaders with opposing political ideologies for one month on a seven point liberal/conservative scale where larger values indicate more conservative opinions about social policy issues for experiments with Democrats (n=697) and Republicans (n=542). Models predict post-treatment liberal/conservative scale score and control for pre-treatment score on this scale as well as twelve other covariates described in the Supplementary Materials. Circles describe unstandardized point estimates, and bars describe 90 and 95 percent confidence intervals. "Respondents Assigned to Treatment" describes the Intent-to-Treat (ITT) effect for Democrats (ITT= -.02, t= -.76, p=.48) and Republicans (ITT=.12, t=2.68, p=.007). "Minimally-Compliant Respondents" describes the Complier Average Causal Effect (CACE) for respondents who followed one of the study's bots for Democrats ($CACE$= -.04, t= -1.07, p=.28) and Republicans ($CACE$=.19, t=5.6, p <.001). "Partially-Compliant Respondents" describes the CACE for respondents who correctly answered at least one question about the content of a bot's tweets during weekly surveys throughout the study period for Democrats ($CACE$=-.05, t= -1.05, p=.29) and Republicans ($CACE$=.31, t=6.30, p <.001), but not all questions. "Fully-Compliant Respondents" describes the CACE for respondents who answered all questions about the content of the bot's tweets correctly for Democrats ($CACE$=-.14, t= -1.05, p=.29) and Republicans ($CACE$=.60, t=4.06, p <.001). Though treated Democrats exhibited slightly more liberal attitudes post-treatment that increase in size with level of compliance, none of these effects were statistically significant. In contrast, treated Republicans exhibited substantially more conservative views post-treatment that increase in size with level of compliance and these effects are highly significant.

on a seven point scale, though our models that estimate the effect of treatment upon fully-compliant Republicans indicates this effect is substantially larger (.60 points), or more than half a standard deviation.

Discussion

Before discussing the implications of these findings, we first place them in context. Though ours is among the largest field experiments conducted on social media to date, the findings above should not be generalized to the entire U.S. population because a majority of Americans do not use Twitter (*21*). Moreover, we did not study people who identify as independents—or those who use Twitter but do so infrequently. Such individuals might exhibit quite different reactions to our intervention. Nevertheless, other studies indicate the type of politically active Americans we studied have an outsized influence on the trajectory of public discussion—particularly as the media itself has come to rely upon Twitter as a source of news and a window into public opinion (*43*). Second, our study offered incentives to encourage respondents to read messages from people with opposing political views, and great care should be taken to generalize the results of this intervention to social media users more broadly, who may simply ignore such counter-attitudinal messages in the absence of such incentives. Finally, our intervention only exposed respondents to high-profile elites with opposing political ideologies. Though our liberal and conservative bots randomly selected messages from across the liberal and conservative spectrum, previous studies indicate such elites are significantly more polarized than the general electorate (*44*). It is thus possible that the backfire effect we identified could be exacerbated by an anti-elite bias, and future studies are needed to examine the effect of inter-group contact with non-elites.

Despite these limitations, our findings have important implications for current debates in sociology, political science, social psychology, communications, and information science. Though we found no evidence that inter-group contact on social media reduces political polarization, our study revealed a significant backfire effect among Republicans who followed our liberal Twitter bot. This is not only the first large-scale field experiment to document the existence of backfire effects that result from repeated interactions between rival groups that occur over an extended time period—and on social media—but also among the first to demonstrate that such effects vary according to partisan identification. To our knowledge, ours is also the first field experiment to disrupt selective exposure to information about politics in a real-world setting via a novel combination of survey research, machine learning, and digital trace data collection. This methodological innovation enabled us to collect information about the nexus of social media and politics with unprecedented granularity while developing new techniques for measuring treatment compliance, mitigating causal interference, and verifying survey responses with behavioral data, as we discuss in our Supplementary Materials. Together, we believe these contributions represent an important advance for the nascent field of computational social science (*45*).

Our research also has urgent implications for policy makers and others who are working to reduce polarization in applied settings. More specifically, our study indicates that well-intentioned attempts to introduce people to opposing political views on social media might not only be ineffective, but counter-productive—particularly if they are initiated by Democrats. Since previous studies have produced substantial evidence that inter-group contact produces compromise and mutual understanding in other contexts, however, future attempts to reduce political polarization on social media will require learning which types of messages, tactics, or issue positions are most likely to create backfire effects and whether others—perhaps delivered by non-elites or in offline settings—might be more effective vehicles to bridge America's partisan divides.

References and Notes

1. P. DiMaggio, J. Evans, B. Bryson, *American Journal of Sociology* **102**, 690 (1996).
2. S. Iyengar, S. J. Westwood, *American Journal of Political Science* **59**, 690 (2015).
3. D. Baldassarri, A. Gelman, *American Journal of Sociology* **114**, 408 (2008).
4. J. Sides, D. J. Hopkins, *Political polarization in American politics* (Bloomsbury Publishing USA, 2015).
5. D. Baldassarri, P. Bearman, *American Sociological Review* **72**, 784 (2007).
6. M. P. Fiorina, S. J. Abrams, *Annual Review of Political Science* **11**, 563 (2008).
7. D. DellaPosta, Shi Yongren, M. Macy, *American Journal of Sociology* **120**, 1473 (2015).
8. C. H. Achen, L. M. Bartels, *Democracy for realists: Why elections do not produce responsive government* (Princeton University Press, 2016).
9. R. S. Erikson, G. C. Wright, J. P. McIver, *Public opinion and policy in the American States* (Cambridge University Press, Cambridge, UK, 1993).
10. J. S. Fishkin, *When the people speak: Deliberative democracy and public consultation* (Oxford University Press, 2011).
11. W. L. Bennett, S. Iyengar, *Journal of Communication* **58**, 707 (2008).
12. C. Sunstein, *Republic.com* (Princeton University Press, Princeton, NJ, 2002).
13. E. Bakshy, S. Messing, L. A. Adamic, *Science (New York, N.Y.)* **348**, 1130 (2015).
14. C. Sunstein, *Echo Chambers, Bush v. Gore , Impeachment, and Beyond* (Princeton University Press, Princeton, NJ, 2001).

15. G. King, B. Schneer, A. White, *Science (New York, N.Y.)* **358**, 776 (2017).

16. J. M. Berry, S. Sobieraj, *The Outrage Industry: Political Opinion Media and the New Incivility* (Oxford University Press, Oxford, UK, 2013).

17. M. Prior, *Annual Review of Political Science* **16**, 101 (2013).

18. E. Pariser, *The Filter Bubble: How the New Personalized Web is Changing What We Read and How We Think* (Penguin, New York, NY, 2011).

19. M. Conover, J. Ratkiewicz, M. Francisco, *ICWSM* **133**, 89 (2011).

20. L. Boxell, M. Gentzkow, J. M. Shapiro, *Proceedings of the National Academy of Sciences* p. 201706588 (2017).

21. A. Perrin, Social Networking Usage: 2005-2015, *Tech. rep.*, Washington, DC (2015).

22. M. Mcpherson, L. Smith-lovin, J. M. Cook, *Annual Review of Sociology* **27**, 415 (2001).

23. A. Edelmann, S. Vaisey, *Poetics* **46**, 22 (2014).

24. D. Lazer, B. Rubineau, C. Chetkovich, N. Katz, M. Neblo, *Political Communication* **27**, 248 (2010).

25. D. Centola, *Science* **334**, 1269 (2011).

26. S. Vaisey, O. Lizardo, *Social Forces* **88**, 1595 (2010).

27. T. F. Pettigrew, L. R. Tropp, *Journal of personality and social psychology* **90**, 751 (2006).

28. R. Huckfeldt, P. E. Johnson, J. Sprague, *Political Disagreement: The Survival of Diverse Opinions within Communication Networks* (Cambridge University Press, Cambridge, UK, 2004).

29. D. C. Mutz, *American Political Science Review* **96**, 111 (2002).

30. R. D. Enos, *Proceedings of the National Academy of Sciences* **111**, 3699 (2014).

31. D. Broockman, J. Kalla, *Science (New York, N.Y.)* **352**, 220 (2016).

32. C. Bail, *Terrified: How Anti-Muslim Fringe Organizations Became Mainstream* (Princeton University Press, 2015).

33. C. G. Lord, L. Ross, M. R. Lepper, *Journal of personality and social psychology* **37**, 2098 (1979).

34. B. Nyhan, J. Reifler, *Political Behavior* **32**, 303 (2010).

35. C. S. Taber, M. Lodge, *American Journal of Political Science* **50**, 755 (2006).

36. T. Wood, E. Porter, *Political Behavior* pp. 1–29 (2016).

37. A. Guess, A. Coppock, *Working Paper* (2017).

38. J. Graham, J. Haidt, B. a. Nosek, *Journal of Personality and Pocial Psychology* **96**, 1029 (2009).

39. J. T. Jost, *et al.*, *Personality and social psychology bulletin* **33**, 989 (2007).

40. M. Grossmann, D. A. Hopkins, *Asymmetric politics: Ideological Republicans and group interest Democrats* (Oxford University Press, 2016).

41. M. Dimock, D. Carroll, *Pew Research Center Report* (2014).

42. P. Barberá, *Political Analysis* **23**, 76 (2014).

43. R. Faris, *et al.*, Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election, *Tech. rep.*, Berkman Klein Center For Internet & Society at Harvard University (2017).

44. A. I. Abramowitz, K. L. Saunders, *The Journal of Politics* **70**, 542 (2008).

45. D. Lazer, *et al.*, *Science (New York, NY)* **323**, 721 (2009).

Supplementary Materials for "Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-scale Field Experiment on Social Media"

Christopher Bail, Lisa Argyle, Taylor Brown, John Bumpus, Haohan Chen, M.B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, Alexander Volfovsky

03/16/17

Contents

1 Introduction

1.1 Replication Materials

This document describes all materials and methods for the article “Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-scale Field Experiment on Social Media,” by Bail et al. All data, code, and the markdown file used to create this report will be available at this link on the Dataverse.

1.2 Pre-registration

The research design and hypotheses described in the main text of our article were pre-registered via the Open Science Framework and can be accessed at this link.

2 Pre-Treatment Survey and Randomization

2.1 Power Analyses

In order to identify a suitable sample size to evaluate our hypotheses, we conducted a literature review for studies about the relationship between exposure to political outgroups and political polarization. The study most similar to our own that we were able to identify was Grönlund et al's (2015) article, “Does Enclave Deliberation Polarize Opinions?,” which appeared in the journal *Political Behavior*. This study presents a field experiment to gauge opinions about immigration in Finland. Out of a total sample of 805 respondents, 366 were assigned to one of two treatment conditions in which they were either asked to deliberate about immigration with a group of people who had heterogeneous views about immigration or one whose members held views that were little or no different than their own. Grönlund et al. report the attitudes of those in the four treatment conditions showed an average change between 0.29 and 1.8 on a scale of 0 to 14 (SD: 2.98)—equal to 0.1 and 0.6 standard deviations.

Figure 1 below reports a power analysis for our study if the effect size were identical to the largest effect reported by Grönlund et al. The red line describes the 80% criterion for probability of sufficient statistical power that is widely employed across the social sciences. According to this calculation, our study would only require approximately 100 people for treatment and control samples. Because our study has two treatment conditions, this would indicate a sample size of 400 is warranted. Yet there are two important differences between this study and our own. First, we planned to expose respondents to out-group Twitter messages for one month, whereas Grönlund et al.'s (2015) study occurred over a single weekend. Second, our study involves virtual contact between respondents and out-groups on a social media site, whereas Grönlund's study involved in-person deliberation.

Though one could argue that the length of our treatment balances out the greater intimacy of Grönlund et al.'s intervention, we believe a more conservative approach is warranted because ours is the first study to examine this process on social media. Compared to in-person deliberation, previous studies indicate online interventions tends to produce smaller attitudinal changes (Luskin, Fishkin, and Hahn 2007). Therefore, we performed an additional power analysis (pictured in Figure 2 below) that estimated ideal sample sizes if the effect we observe is half the size of the largest effect from Grönlund et al.'s (2015) study. This illustration indicated that approximately 300 respondents are required per treatment/control comparison (or 1,200 respondents overall between the Democratic and Republican experiments).

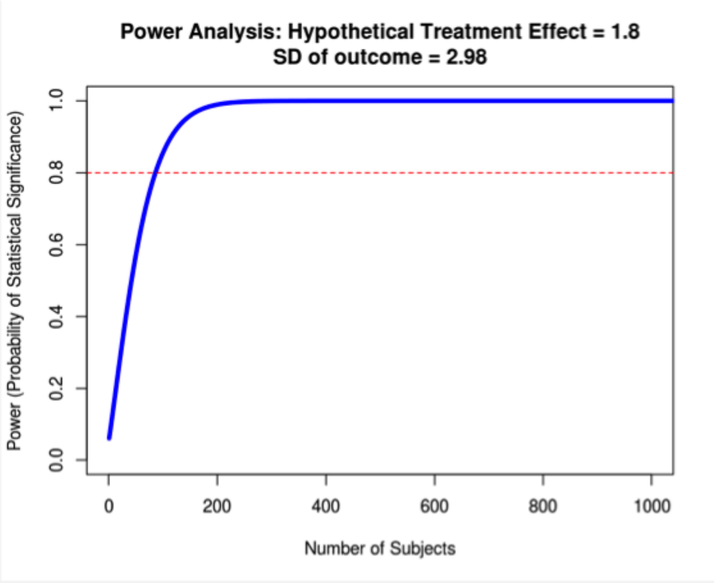


Figure 1: Power Analysis #1.

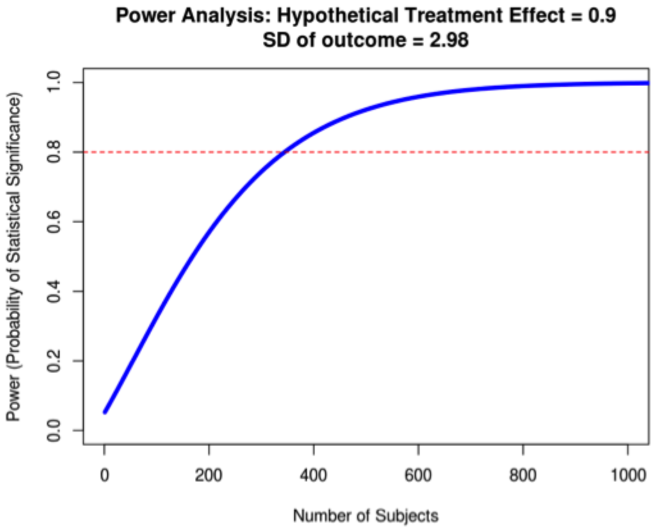


Figure 2: Power analysis #2.

2.2 Survey Recruitment Process

We hired YouGov—one of the largest and most reputable survey firms in the United States—to recruit at least 1,200 self-identified Republicans and Democrats over age 17 who visit Twitter at least three times each week to complete five surveys between mid-October 2017 and mid-November 2017. A more detailed description of YouGov and its web panelists is available here. Figure 3 provides a detailed description of the recruitment process for our pre-treatment survey, which was fielded between October 10th and October 19th, 2017. YouGov invited 10,634 members of its U.S. panel to participate in our study using U.S. census sampling frames. Of these, 5,520 did not respond, and 5,114 accepted the invitation, for an initial cooperation rate of 48% (AAPOR RR3 = 42.7%). These individuals were then asked several screening questions. First, they were asked about their party identification using the following question: "Generally speaking, do you think of yourself as a [Democrat/Republican/Independent/Other/Not Sure]." Respondents who did not respond with either "Democrat" or "Republican" were screened out, and remaining respondents were asked the following question, which was used to identify the treatment blocks described in further detail below: "Would you call yourself a strong Democrat/Republican or a not very strong Democrat/Republican?" These two questions have been widely employed to measure party attachment in the American National Election Study and many other surveys. Third, respondents were asked if they "visit Twitter at least three times a week in order to read messages from other Twitter accounts," and screened out if they answered negatively.

A total of 2,539 people were deemed eligible according to these two initial eligibility criteria, and were subsequently read an informed consent dialogue and offered the equivalent of $11 via YouGov's "points" system, which allows respondents to redeem points for items such as Amazon gift cards, to share their Twitter handle, or Twitter ID, in order that it may be linked to their survey responses. 1,754 agreed and completed the entire pre-treatment survey. 500 respondents began—but did not complete—the pre-treatment survey, and 285 respondents refused to complete the pre-treatment survey. Of the 1,754 respondents that completed the pre-treatment survey, 102 were removed by YouGov's quality algorithm, which eliminates respondents who complete the survey within a time frame that is deemed impossible by the algorithm. This resulted in an initial sample of 1,652 respondents.

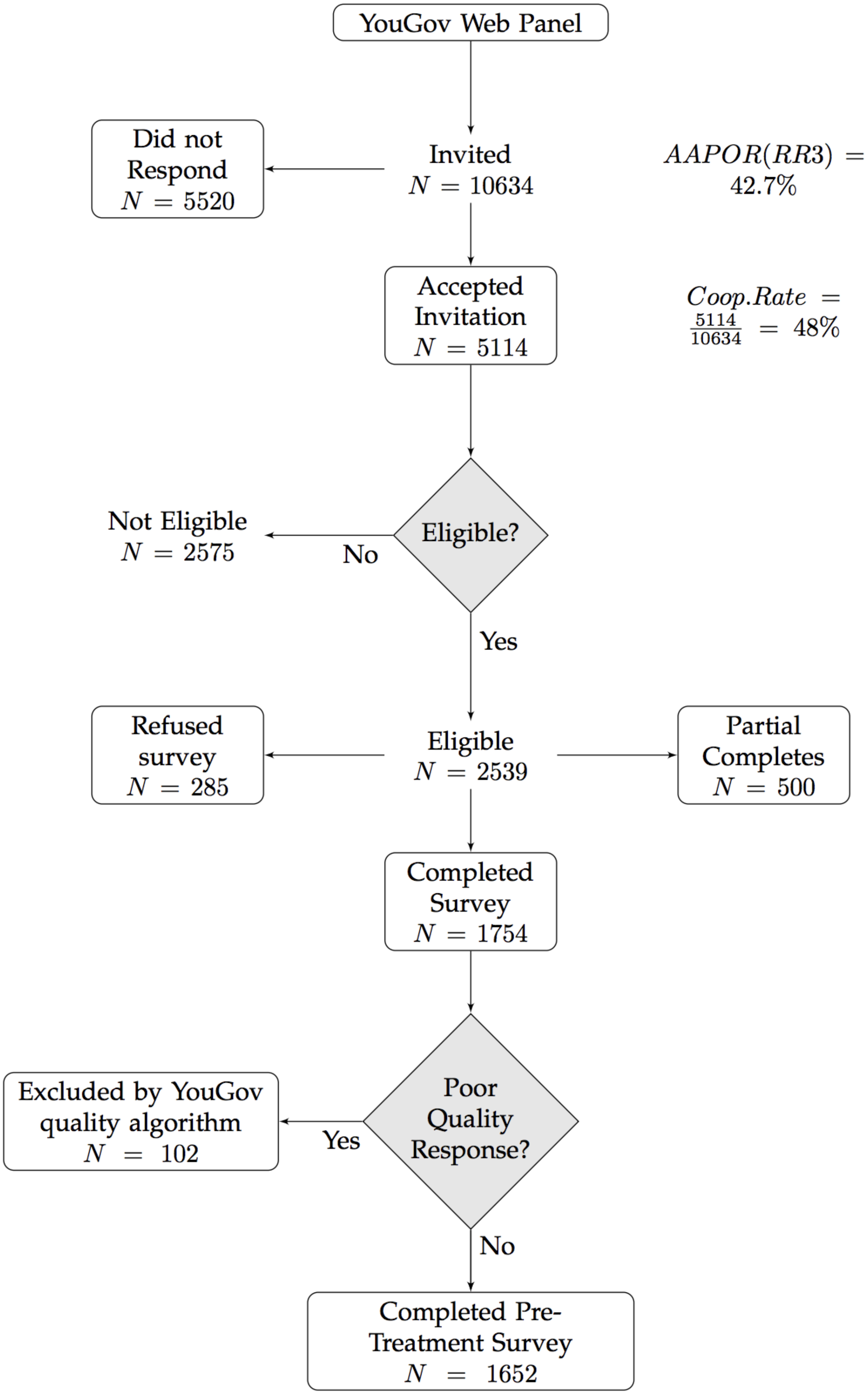

YouGov Web Panel
Did not Respond $N = 5520$
Invited $N = 10634$
$AAPOR(RR3) = 42.7\%$
Accepted Invitation $N = 5114$
$Coop.Rate = \frac{5114}{10634} = 48\%$
Not Eligible $N = 2575$
No
Eligible?
Yes
Refused survey $N = 285$
Eligible $N = 2539$
Partial Completes $N = 500$
Completed Survey $N = 1754$
Excluded by YouGov quality algorithm $N = 102$
Yes
Poor Quality Response?
No
Completed Pre-Treatment Survey $N = 1652$

Figure 3: Recruitment Process for Pre-Treatment Survey

2.3 Respondents Eliminated Before Treatment Assignment

136 of the 1,652 respondents who completed the pre-treatment survey were excluded from subsequent analyses because they did not present a valid Twitter handle or username that could be accessed via Twitter's Application Programming Interface. Forty-five respondents were excluded because they provided poor quality data, indicated by providing the same answer to ten consecutive questions that were randomized according to whether the respondent was asked to agree with a liberal or conservative-leaning statement (an additional twelve respondents were later excluded because they did the same during the final post-treatment survey).

A major advantage of our research design is that we were able to cross-validate survey responses with behavioral and demographic information available from the Twitter profiles and messages of respondents to our pre-treatment survey. Forty-four respondents were excluded because they did not follow any accounts on Twitter, and therefore did not satisfy our screening criterion that participants be active Twitter users who "regularly log on to read messages from other Twitter accounts." We elected to exclude any respondent for whom demographic information in the survey conflicted with at least two demographic variables that were observable on the respondent's Twitter page (age, gender, race, and geographic location). Some of these respondents were excluded because of the aforementioned exclusion rules, but an additional 74 respondents were dropped because they provided highly inconsistent information in the survey and in their Twitter profile. 4 additional respondents were excluded because we suspected they provided an account of a famous person instead of their own Twitter account. We operationalized fame as having more than 50,000 followers. Because it was theoretically possible that a respondent in our study could have a large number of followers, we cross-referenced demographic information from the Twitter account in question with that reported in our survey and identified significant discrepancies which further increased our confidence that these responses were non-valid.

2.4 Causal Interference

Yet another advantage of our research design is that we were able to collect social network data from each respondent's Twitter account in order to mitigate the risk of causal interference in our survey population. For example, respondents in our control condition could receive partial treatment if they follow a respondent in the treatment condition who retweeted—or otherwise engaged with—a message produced by one of the bots created by our research team. After removing respondents who were excluded for reasons described in the previous section of this document, we identified 136 respondents in our sample who followed—or were followed by—at least one other respondent in our study. As Figure 4 shows, 90 of these people were part of network components that included at least three people.

We excluded all respondents that were part of such components from the current study, but treated some of them as part of a separate study designed to gauge opinion-leader dynamics that we will report at a later date. Of the remaining 46 people who were connected to only one other person in the survey population, we randomly dropped one respondent within each dyad. Figure 5 plots these network relationships without isolates to further illustrate our decisions.

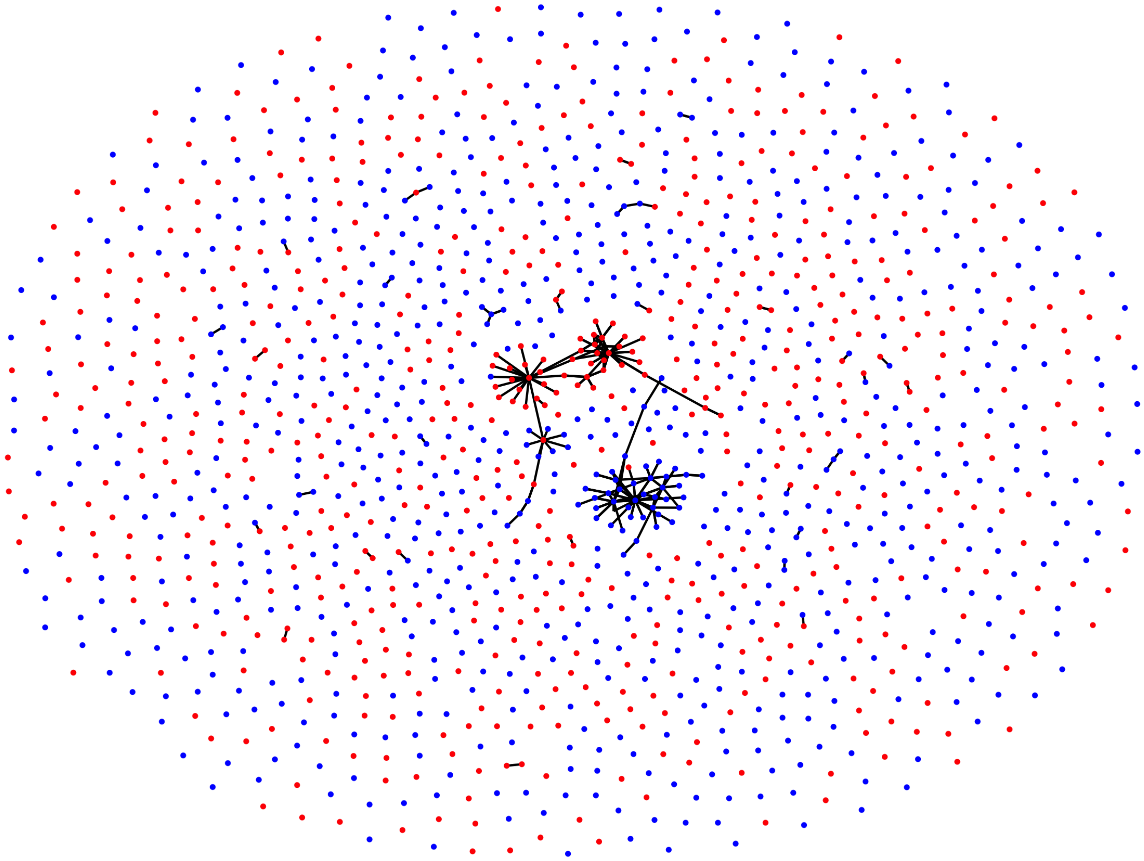


Figure 4: Network Diagram Describing Twitter Connections between Respondents in Initial Pre-Treatment Survey

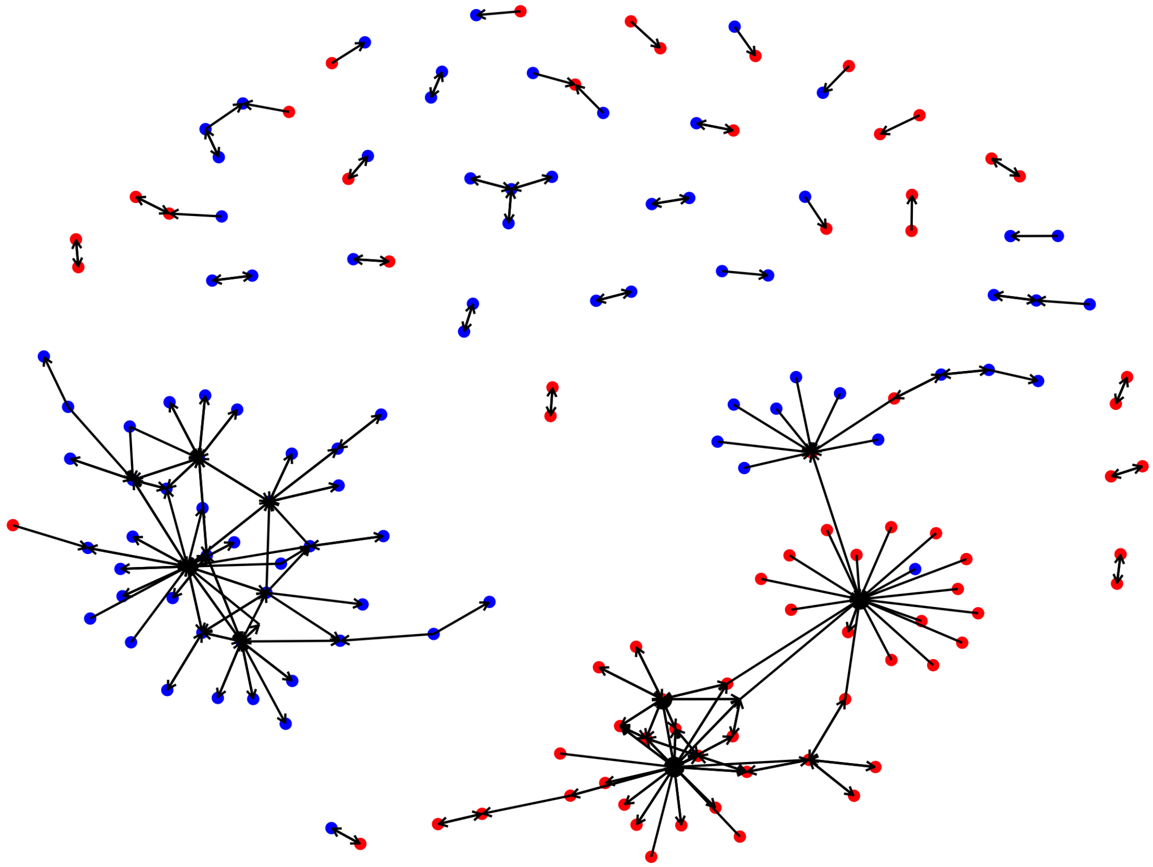
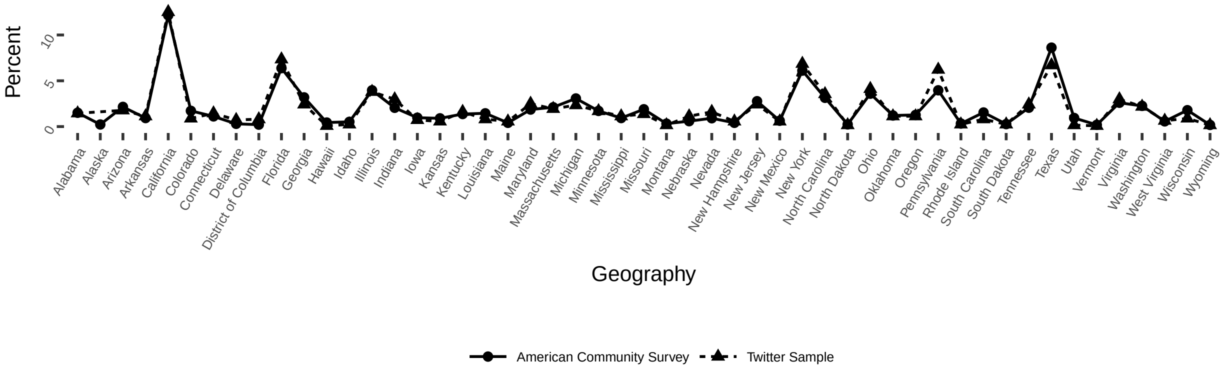
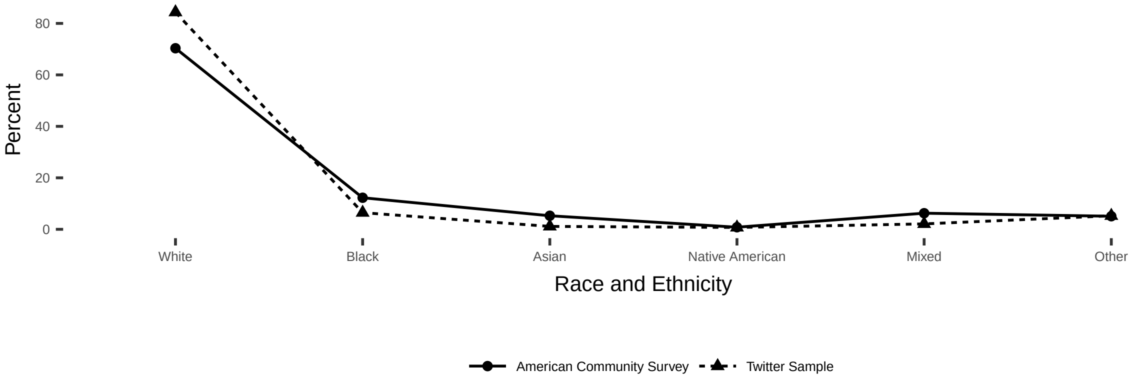


Figure 5: Network Diagram Describing Twitter Connections between Respondents in Initial Pre-Treatment Survey (without isolates)

2.5 Descriptive Characteristics of Final Study Population

Our final sample included 1,220 respondents (691 self-identified Democrats, and 529 self-identified Republicans). The figures below compare our sample to data from the 2016 American Community Survey, which are available here. Data on state populations for 2016 were collected from the U.S. Census, and are available here. As these figures show, our sample closely approximates the adult population distribution across U.S. states, races and ethnicities, and gender, as the following figures demonstrate.





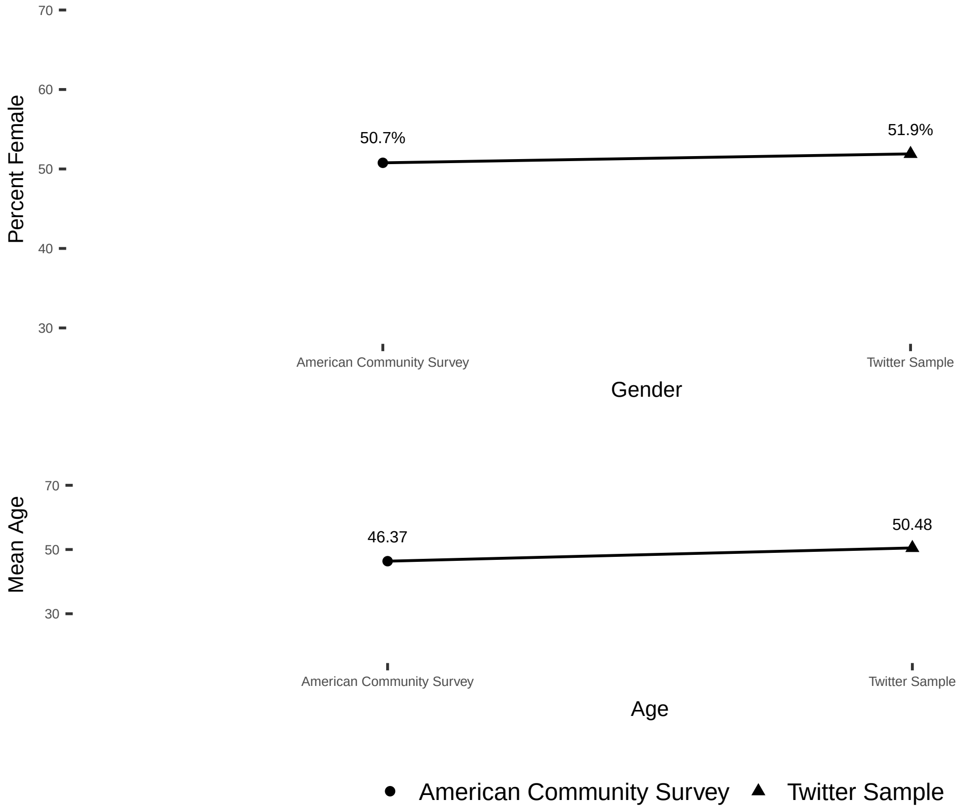


Figure 6: Comparison of Demographic Characteristics of Respondents in Study Sample to U.S. Census/American Community Survey

2.6 Block Randomization

In order to ensure sufficient statistical power we employed a block randomized design. We identified two strong covariates of political polarization based upon a comprehensive review of the literature. These included a) level of attachment to political party; and, b) overall level of interest in news and current events. We also stratified by frequency of Twitter usage, because we reasoned that more frequent Twitter users would be exposed to more of the messages produced by our study's Twitter bots. One week after the pre-treatment survey, all respondents who provided valid responses were randomized to treatment using the aforementioned design.

To measure respondents' attachment to their party, we employed the aforementioned question that is used widely within the literature and prominent studies of the political attitudes of the American public. To measure overall level of news interest, respondents were asked the following question: "Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs . . . " (Most of the time, Some of the time, Only Now and Then, Hardly at all, Don't Know). The wording of this question was adopted verbatim from the American National Election Study.

To measure frequency of Twitter use, we asked respondents the following question "How often do you visit Twitter to read posts from other accounts?" We then created a binary variable that describes whether people visited Twitter multiple times every day, or less than two times each day. We also considered using the number of accounts respondents followed on Twitter as a criterion for overall frequency of Twitter activity, but this measure was highly correlated with the self-reported measure in our survey just described and does not necessarily reflect the regularity with which a user visits Twitter.

2.7 Covariate Balance Check

The table below reports the results of t tests that show no significant difference in covariates between respondents who were assigned to the treatment condition and those who were not.

Table 1: Covariate Balance

Covariate	p-value
Geography (Northeast)	0.14
Geography (South)	0.42
Geography (North Central)	0.78
Geography (West)	0.45
Race	0.46
Gender	0.14
Age	0.08

Because we ran separate experiments for Democratic and Republican respondents, we also ran separate covariate balance checks by party identification. These tests revealed balance on almost all of our control variables, apart from the dummy variable describing respondent location in the Northeastern United States (for Republicans), and gender (for Democrats). Although neither of these variables have been shown to have a particularly strong relationship with *change* in partisan identification by previous studies, we control for them in the models described below, and ran separate models that exclude these variables which produced results that are nearly identical to those which we report in the main text of our article.

Table 2: Covariate Balance, Republican Experiment

Covariate	p-value
Geography (Northeast)	0.02
Geography (South)	0.79
Geography (North Central)	0.52
Geography (West)	0.21
Race	0.98
Gender	0.99
Age	0.08

Table 3: Covariate Balance, Democrat Experiment

Covariate	p-value
Geography (Northeast)	1
Geography (South)	0.39
Geography (North Central)	0.37
Geography (West)	0.94
Race	0.38
Gender	0.05
Age	0.42

3 Treatment Delivery and Compliance

Respondents assigned to the treatment condition were invited to follow the bots on October 21, 2017, roughly one week after they were recruited to complete the pre-treatment survey. The one week buffer between the initial survey and treatment was intended to decrease the likelihood that respondents would become aware of the purpose of the experiment (see additional discussion of experiment effects below). The invitation to those in the treatment condition was as follows: "Recently you completed a survey for YouGov about how often you access Twitter and other news sources online. You have been randomly selected for an opportunity to receive up to 25,000 points for completing additional tasks related to that survey. Participation in this portion of the study will involve following a Twitter account created by the sponsor of this study for one month." Respondents were then redirected to another page that included a link to follow either the study's liberal or conservative bot, depending upon their self-reported party identification. This page informed them they would earn the equivalent of $11 for following the bot and up to an additional $18 for successfully answering questions about the content of the messages retweeted by the bot during surveys that would follow each week. Each bot was given a non-descript name that did not prime the political ideology of opinion leaders that it retweeted. We are unable to report the names here because data collection continues for follow-up research and disclosure of the twitter handles could be used to identify respondents in the study who engaged with the bot by commenting on, liking, or retweeting its retweets. For the first few days in which respondents began following the bots, only pictures of nature landscapes were retweeted in order to further mask the purpose of the study.

3.1 Ethics and Protection of Human Subjects

Our research was approved by the Institutional Review Boards at Duke University and New York University. All respondents digitally signed an informed consent dialogue before they participated in our research. Because open-ended questions in our pilot study indicated Republican respondents might have anti-intellectual sentiment that could create measurement error, the informed consent dialogue did not state that the research was being conducted by academic researchers, though the name of the first author's university was listed alongside instructions about how to contact the Institutional Review Board with complaints about the research. No such complaints were received.

Though most Twitter data is publicly available for academic research, our study links such data to confidential survey data. Because such data are highly sensitive, we do not publicly release the names, twitter handles, or numeric ids of any respondents in our study. Nor do we make the content of their tweets, the names of the people they follow, or the names of the people who follow them publicly available. Instead, the public release of our data will only contain each respondent's a) number of twitter followers, and b) the number of people they follow. Each member of the research team who analyzed the non-public Twitter data signed a confidentiality pledge in accordance with evolving standards in the field of social media research.

3.2 Algorithmic Confounding

A unique property of social media research, as discussed by Lazer et al (2014), is algorithmic confounding—or the possibility that software that governs users' experience on social media websites can shape research findings. In our case, we were concerned that Twitter's "timeline" algorithm. This algorithm sorts the order in which messages appear in a user's twitter feed, which we feared could shape exposure to our bots, since Twitter's algorithm prioritizes accounts with which a user regularly engages (either by retweeting messages from such accounts, commenting on messages produced by such accounts, or "liking" messages produced by such accounts).

To mitigate algorithmic confounding, we asked all respondents to disable Twitter's timeline algorithm in order to ensure that they viewed tweets from their bot, and were thus able to answer questions about the content of its tweets. We provided step-by-step instructions about how to disable the algorithm in our recruitment dialogue.

We also took steps to mitigate bias from Twitter's "recommender" algorithm. Our concern was that people who follow one of our study's bots may subequently receive recommendations to follow similar types of people (i.e. out-partisans), which could effectively strengthen our treatment (if the user acts upon such recommendations). Though Twitter does not make details about its recommender algorithm publicly available, Gupta et al (2013), who were involved in the design of the algorithm, indicate it is makes recommendations using second-order network relationships—or who the person who is followed, follows. Because our bots did not follow any other Twitter accounts, we believe the likelihood of bias created by the recommender algorithm was very low.

3.3 Measuring Compliance

We measured compliance with treatment assignment in several ways described in our pre-registration statement. First, we wrote code to monitor whether respondents followed one of our bots for the entire study period, and also to collect supplemental social network data that we will analyze in a subsequent study. Because it is possible that some respondents followed us but subsequently "muted" our bot—or changed the settings for their Twitter account so that they continue to follow one of our bots but do not receive its messages within their timeline—we employed deception. In our recruitment dialogue, we informed respondents that the sponsor of the research would monitor whether they muted the account using a computer program, even though it is not possible to do so. This deception was approved by the Institutional Review Board that monitored our research and subjects were debriefed about this deception following the conclusion of our study.

Because we could not be certain that respondents who followed one of our study's bots were consistently exposed to the messages it retweeted, we took two additional steps. First, we conducted weekly surveys of respondents which asked them to answer substantive questions about the content of our Twitter bots' messages during the previous week. For example, we asked respondents to provide information from one tweet during the previous 72 hours that could not be easily searchable by using a browser's "search" function. We carefully designed these questions so as not to favor those with higher political knowledge or prime partisan sentiment. We are unable to list the exact wording of these questions because such information could be used to identify the name of our Twitter bots, and hence, the names of those in the study who interacted with its messages throughout the study period.

In addition to these substantive questions, the Twitter bots retweeted a picture of an animal twice each day for a week. The animal pictures were deleted from the bot accounts immediately prior to sending the weekly compliance check survey, in which respondents were asked to identify the animal picture that was tweeted in the prior week. We developed this additional compliance measure in order to determine whether respondents were exposed to messages produced by the bots in real time, or if they only read its messages at the time of the compliance survey. As we mention in the main text of our article, 40.14 percent of respondents who followed the bots successfully answered all weekly substantive questions about the bots' tweets, and 32.48 percent of respondents successfully identified all animal pictures. The three weekly surveys which asked respondents these questions—as well as the substantive questions described above– were conducted between October 27-29th, November 3-5, and November 10-13th. The final post-treatment survey, which contained the exact same questions as the pre-treatment survey, was administered between November 23rd to November 27th.

The code below calculates the compliance rate by party in three ways. First, we calculate the percentage of people who accepted our invitation to follow one of the study's bots. Second we create a six-point compliance scale that describes the number of questions that respondents were able to answer correctly during the three "compliance check" surveys administered each week during the study period.

```r
#calculate compliance rate by party identification
democrats<-twitter_data[twitter_data$party_id_wave_1==1,]
nrow(democrats[democrats$bot_followers==1,]) / nrow(democrats[democrats$treat==1,])
republicans<-twitter_data[twitter_data$party_id_wave_1==2,]
nrow(republicans[republicans$bot_followers==1,]) / nrow(republicans[republicans$treat==1,])
```

```r
#construct continuous compliance measure
twitter_data$complier_scale<-0
twitter_data$complier_scale <-
  rowSums(twitter_data[,c("substantive_question_correct_wave_2",
  "substantive_question_correct_wave_3",
  "substantive_question_correct_wave_4",
  "animal_correct_wave_2",
  "animal_correct_wave_3",
  "animal_correct_wave_4")], na.rm=TRUE)


#construct partial complier dummy
twitter_data$half_complier<-0
twitter_data$half_complier[twitter_data$complier_scale>0&
                           twitter_data$complier_scale<6]<-1

#construct full complier dummy
twitter_data$perfect_complier<-0
twitter_data$perfect_complier[twitter_data$complier_scale==6]<-1


partial_compliance_rate<-nrow(twitter_data[(twitter_data$half_complier==1),])/
       nrow(twitter_data[(twitter_data$half_complier==0),])

full_compliance_rate<-nrow(twitter_data[(twitter_data$perfect_complier==1),])/
       nrow(twitter_data[(twitter_data$perfect_complier==0),])
```

3.4 Design of Twitter Bots

The two Twitter bots we created for this study were designed as follows. First, we built upon Barbera et al.'s (2015) ideological scoring method for Twitter users. We began by collecting the Twitter IDs, or "handles," for all presidential candidates and members of the House and Senate as of August, 5 2017. We then scraped the names of all people who these elected officials follow from Twitter's Application Programming Interface, which yielded a total sample of 636,737 Twitter accounts. Next, we eliminated all those who were not followed by at least 15 of the aforementioned elected officials. We then conducted a correspondence analysis on the resultant adjacency matrix, and used the first principal component to create a liberal/conservative score for all of those in this "opinion leader" network. We binned this scale into seven quantiles, and dropped those in the fourth, centrist, quantile. The liberal bot randomly retweeted messages from opinion leaders in the first, second, and third quantile produced during the preceeding 24 hours, and the conservative bot randomly retweeted messages from opinion leaders in the fifth, six, and seventh quartiles during the preceeding 24 hours.

We took several additional steps to improve the ideological scores we used to create our bots. First, we eliminated all U.S. government agencies, since most of these retweeted non-partisan messages that would dilute our treatment. Second, we eliminated all accounts that were administered by for-profit U.S. corporations, though we did not eliminate non-profit organizations, think tanks, or other nonprofit groups. Third, we eliminated a small number of accounts that were controlled by elected officials outside the United States.

Despite these steps, pilot analyses of the ideological continuum consistently identified a small number of elected officials who were misclassified according to our measure. Each of these individuals were very high profile opinion leaders such as Mitch McConnell and John McCain, who have very large followings that include a large number of non-Republicans, which made them centrists instead of conservatives in our original analysis. We thus reclassified the small number of elected officials who were mistakenly identified by assigning them a random ideological score between the first and second quantile of opinion leaders that defined their party using the first principal component measure described above.

The liberal and conservative Twitter bots created by our research team were both hosted on an Amazon EC2 server. Every hour, our program randomly drew a message produced by an elected official or opinion leader from the previous 24 hours from one of the two samples. During three of the four weeks during the study period, the bot retweeted a different animal picture at two random times each day, as we described above.

To further illustrate the types of Twitter accounts retweeted by the study's bots, Table 4 lists the words that appear most frequently in the "about me," or biographical section, of the Twitter accounts retweeted by both the liberal and conservative bots. Similarly, Table 5 lists the top words that appeared in messages retweeted by the two bots during the study period.

Table 4: Top Words from Biographies of Twitter Accounts Retweeted by Bots

Liberal Accounts	Freq.	Conservative Accounts	Freq.
news	103	news	74
reporter	54	district	60
politics	49	u.s	53
political	48	official	41
u.s	43	conservative	37
correspondent	42	congressional	33
house	42	proudly	33
editor	41	serving	31
national	36	twitter	30
cnn	34	chairman	29
official	34	senator	29
washington	34	policy	28
district	33	account	26
politico	33	house	26
covering	30	author	25
tweets	29	husband	25
author	27	follow	24
twitter	26	people	24
congressional	25	fox	23
endorsement	25	host	23
host	25	father	22
senior	25	congress	21
representing	24	represent	21
account	23	representing	21
policy	23	american	20
alum	21	media	20
follow	21	editor	19
congress	20	national	18
health	20	politics	18
email	19	proud	17
people	19	republican	17
world	19	tweets	17
writer	19	united	17
american	18	political	16
chief	18	president	16
breaking	17	america	15
tips	17	business	15
white	17	committee	15
america	16	senate	15
security	16	congressman	14
times	16	contributor	14
analysis	15	life	14
endorsements	15	organization	14
msnbc	15	public	14
public	15	research	14

Table 5: Top Words from Messages Retweeted by Bots

Liberal Bot’s Tweets	Freq.	Conservative Bot’s Tweets	Freq.
trump	55	tax	45
tax	26	house	28
people	23	trump	24
it’s	20	act	21
time	18	jobs	21
health	13	people	19
plan	13	time	19
day	12	day	18
president	12	reform	16
story	12	support	15
week	12	taxreform	15
campaign	11	news	14
change	11	veterans	14
twitter	11	american	13
climate	10	discuss	13
families	10	read	13
gop	10	service	13
live	10	bill	12
talking	10	passed	12
fight	9	http	11
house	9	senate	11
tomorrow	9	clinton	10
veterans	9	cuts	10
watch	9	families	10
attack	8	join	10
bill	8	live	10
news	8	morning	10
read	8	u.s	10
service	8	week	10
sexual	8	forward	9
talk	8	happy	9
taxes	8	hillary	9
vote	8	icymi	9
american	7	potus	9
care	7	russia	9
election	7	watch	9
hearing	7	america	8
local	7	create	8
pay	7	dossier	8
russian	7	free	8
tune	7	gop	8
tweet	7	honor	8
tweets	7	it’s	8
colleagues	6	listen	8
fire	6	maga	8

4 Outcome Measure and Controls

4.1 Creating Outcome Index

Our study examined a range of political attitudes, but in this article we focus upon shifts in what is often called "ideological polarization," or differences in attitudes about policy issues that consistently divide Democrats and Republicans such as inequality, race, and immigration. Though ideological scoring of roll-call voting has been the subject of extended analysis for some time, indices of liberalism vs. conservatism among the broader public are fewer (Bafumi and Herron 2010; Jessee 2012; Tausanovitch and Warshaw 2013; Hill and Tausanovitch 2015). We employed a variation of the "ideological consistency scale" developed by previous studies because it measures liberal vs. conservative opinions via a battery of questions in order to minimize the measurement error that might occur on a single survey item (Dimock and Carroll 2014). The scale, which asks respondents to agree or disagree with a series of twenty statements about social policies worded to favor either liberal or conservative views, was previously included in sixteen nationally representative surveys. We make two important modifications to this scale. Instead of a binary choice between liberal and conservative options for each policy statement, we use a seven-point response scale, since allowing respondents to indicate strength or extremity of opinion provides a more accurate measure of ideological polarization (Fiorina, Abrams, and Pope 2006; Hill and Tausanovitch 2015). Second, instead of asking respondents to read twenty questions, we randomly selected five liberal versions of each policy statement and five conservative versions. These modifications were important because we did not expect our intervention would completely change people's partisan positions (from Democrats to Republicans), and also because the format just described minimizes cognitive load.

Our survey asked respondents to agree or disagree with the following statements on a seven point scale from "strongly disagree" to "strongly agree."

1) "Stricter environmental laws and regulations cost too many jobs and hurt the economy."
2) "Government regulation of business is necessary to protect the public interest."
3) "Poor people today have it easy because they can get government benefits without doing anything in return."
4) "Immigrants today strengthen our country because of their hard work and talents."
5) "Government is almost always wasteful and inefficient."
6) "The best way to ensure peace is through military strength."
7) "Racial discrimination is the main reason why many black people can't get ahead these days."
8) "The government today can't afford to do much more to help the needy."
9) "Business corporations make too much profit."
10) "Homosexuality should be accepted by society."

As we mentioned, half of these statements are worded in a manner that is designed to appeal to liberals (#2,#4,#7,#9,#10), and the other half are intended to appeal to conservatives (#1,#3,#5,#6,#8). Question order was randomized in both the pre and post-treatment surveys.

The code below was used to create our outcome measure. Liberal questions were reverse-coded such that negative values on our outcome indicate respondents becoming more liberal and positive values indicate respondents becoming more conservative. We calculate the mean score on this ten-item index for the pre and post-treatment survey, and our models predict the post-treatment scale score, controlling for the pre-treatment scale score.

```
#invert questions that prime liberal values
twitter_data$government_should_regulate_businesses_wave_1<-
```

```
  8-twitter_data$government_should_regulate_businesses_wave_1
twitter_data$racial_discrimination_hurts_black_people_wave_1<-
  8-twitter_data$racial_discrimination_hurts_black_people_wave_1
twitter_data$immigrants_strengthen_country_wave_1<-
  8-twitter_data$immigrants_strengthen_country_wave_1
twitter_data$corporations_make_too_much_profit_wave_1<-
  8-twitter_data$corporations_make_too_much_profit_wave_1
twitter_data$homosexuality_should_be_accepted_wave_1<-
  8-twitter_data$homosexuality_should_be_accepted_wave_1
twitter_data$government_should_regulate_businesses_wave_5<-
  8-twitter_data$government_should_regulate_businesses_wave_5
twitter_data$racial_discrimination_hurts_black_people_wave_5<-
  8-twitter_data$racial_discrimination_hurts_black_people_wave_5
twitter_data$immigrants_strengthen_country_wave_5<-
  8-twitter_data$immigrants_strengthen_country_wave_5
twitter_data$corporations_make_too_much_profit_wave_5<-
  8-twitter_data$corporations_make_too_much_profit_wave_5
twitter_data$homosexuality_should_be_accepted_wave_5<-
  8-twitter_data$homosexuality_should_be_accepted_wave_5

#calculate chronbach's alpha

alpha_calc<-twitter_data[,c(
          "government_should_regulate_businesses_wave_1",
          "racial_discrimination_hurts_black_people_wave_1",
          "immigrants_strengthen_country_wave_1",
          "corporations_make_too_much_profit_wave_1",
          "homosexuality_should_be_accepted_wave_1",
          "government_wasteful_inefficient_wave_1",
          "poor_people_have_it_easy_wave_1",
          "government_cannot_afford_to_help_needy_wave_1",
          "best_way_peace_military_strength_wave_1",
          "stricter_environmental_laws_damaging_wave_1")]
library(psych)
psych::alpha(alpha_calc)

#create average score by wave
twitter_data$substantive_ideology_scale_wave_1<-rowMeans(twitter_data[,c(
          "government_should_regulate_businesses_wave_1",
          "racial_discrimination_hurts_black_people_wave_1",
          "immigrants_strengthen_country_wave_1",
          "corporations_make_too_much_profit_wave_1",
          "homosexuality_should_be_accepted_wave_1",
          "government_wasteful_inefficient_wave_1",
          "poor_people_have_it_easy_wave_1",
          "government_cannot_afford_to_help_needy_wave_1",
          "best_way_peace_military_strength_wave_1",
          "stricter_environmental_laws_damaging_wave_1")], na.rm=TRUE)


twitter_data$substantive_ideology_scale_wave_5<-rowMeans(twitter_data[,c(
          "government_should_regulate_businesses_wave_5",
          "racial_discrimination_hurts_black_people_wave_5",
          "immigrants_strengthen_country_wave_5",
          "corporations_make_too_much_profit_wave_5",
          "homosexuality_should_be_accepted_wave_5",
          "government_wasteful_inefficient_wave_5",
          "poor_people_have_it_easy_wave_5",
          "government_cannot_afford_to_help_needy_wave_5",
          "best_way_peace_military_strength_wave_5",
          "stricter_environmental_laws_damaging_wave_5")], na.rm=TRUE)
```

4.2 Control Variables

The models described below include a variety of control variables collected from our pre-treatment survey, Twitter's Application Programming Interface, and YouGov. We also ran separate models without controls, which yielded very similar results (Republicans: t=2.74, p<.006, Democrats: t= -1.54, p<.12). We obtained standard demographic variables about all respondents (age, income, education, gender, race, and geographic region) from YouGov. We also created variables designed to measure the strength of respondents' echo chambers pre-treatment. We asked respondents a battery of questions about their media consumption practices, and requested they list the top three media sources they consume most frequently in order to determine the amount of ideological bias in their media diet pre-treatment. Unfortunately, more than 25% of respodents did not provide the names of media sources for which we could identify ideological leaning, so we were ultimately unable to include this variable in our analyses. Fortunately, we also calculated the percentage of people who the respondent follows on Twitter who share their party identification pre-treatment using the network-based ideological scoring method for Twitter user's described above. This measures is highly correlated with the aforementioned media consumption measure, so we include the Twitter-based metric to measure the strength of respondents' echo chambers pre-treatment. Our pre-treatment survey also asked respondents to estimate the percentage of people in their offline networks who share their party identification in order to further capture the strength of ideological bias within their offline social networks, which have been shown to have higher ideological bias than online networks (Gentzkow and Shapiro 2011). We include this continuous measure in all of our models as well. We did not detect significant multicollinearity between these two variables.

The code below recodes and subsets the control variables for analyses that we conduct below. The "bin_maker" variable describes the randomization blocks described below.

```
#subset control variables and variable used for block randomization (bin_maker)
control_variables<-twitter_data[,c(
    "percent_co_party",
    "political_wave_1",
    "freq_twitter_wave_1",
    "friends_count_wave_1",
    "strong_partisan",
    "birth_year",
    "family_income",
    "education",
    "gender",
    "ideo_homogeneity_offline",
    "northeast",
    "north_central",
    "south",
    "west",
    "caseid",
    "bin_maker")]
```

4.3 Missing Data

We employed multiple imputation to address a small amount of missing data in our pre-treatment survey–particularly the variables that describe respondents' income (7% missing) as well as their estimate of the ideological composition of their offline networks (1% missing). We did not impute missing responses for the second wave outcome measure.

```
#examine missing data in first wave

library(Amelia)
missmap(control_variables)

#impute missing data from first wave
forimputation<-cbind(twitter_data$substantive_ideology_scale_wave_1, control_variables)
colnames(forimputation)[colnames(forimputation)==
              "twitter_data$substantive_ideology_scale_wave_1"]<-
                      "substantive_ideology_scale_wave_1"

library(mice)
#prepare variables for imputation
forimputation$caseid<-as.character(forimputation$caseid)
#take log of variables with heavy skew
forimputation$percent_co_party<-log(forimputation$percent_co_party+1)
forimputation$friends_count_wave_1<-log(forimputation$friends_count_wave_1+1)

#impute
imputed_data <- mice(forimputation,m=15,seed=352,
                     exclude=c("caseid","bin_maker"))
imputed_data <- complete(imputed_data,action=15)

#reassemble dataaset with additional variables we need for subsequent analysis
to_bind<-twitter_data[,c("treat",
                  "perfect_complier",
                  "half_complier",
                  "bot_followers",
                  "party_id_wave_1",
                  "party_strength_wave_1",
                  "substantive_ideology_scale_wave_5",
                  "endtime_wave_5")]

final_data<- cbind(to_bind, imputed_data)


save(final_data, file="Final Data for Models.Rdata")
```

5 Calculating Treatment Effects

To evaluate the effect of our intervention, we calculated both Intent-to-Treat (ITT) effects as well as Complier Average Causal Effects (CACE) that account for level of treatment compliance for experiments with both Democrats and Republicans. These models predict the post-treatment score on our ten-item liberal/conservative scale controlling for respondents' pre-treatment score on this scale, twelve additional covariates (described above), as well as a factor variable (bin_maker) that describes our treatment blocks.

5.1 Intent-to-Treat Effects

```
#Subset Republicans and Democrats and Drop Missing Data from Post-Treatment Survey

republicans <- final_data[final_data$party_id_wave_1==2,]
republicans<-republicans[complete.cases(republicans),]
democrats <- final_data[final_data$party_id_wave_1==1,]
democrats<-democrats[complete.cases(democrats),]

#Republicans

republican_ITT_model<-lm(substantive_ideology_scale_wave_5~
                      #treatment assignment variable
                      treat+
                      #pre-treatment ideology score
                      substantive_ideology_scale_wave_1+
                      #% of people followed on Twitter from same party
                      percent_co_party+
                      #% of people in offline networks from same party
                      ideo_homogeneity_offline+
                      #total number of people followed pre-treatment
                      friends_count_wave_1+
                      #demographics
                      birth_year +
                      family_income+
                      education+
                      gender+
                      northeast+
                      north_central+
                      south+
                      #factor variable used to create treatment blocks
                      as.factor(bin_maker),
                      data=republicans)

coefficients<-data.frame(summary(republican_ITT_model, cluster="bin_maker")$coefficients[2:13,])
library(pander)
panderOptions('digits',3)
panderOptions('table.split.table', 300)
set.caption("Intent-to-Treat Model (Republicans)")
pander(coefficients)
```

Table 6: Intent-to-Treat Model (Republicans)

	Estimate	Std..Error	t.value	Pr...t..
treat	0.12	0.0448	2.68	0.00761
substantive__ideology__scale__wave__1	0.819	0.0277	29.6	1.4e-104
percent__co__party	0.227	0.11	2.06	0.0401
ideo__homogeneity__offline	0.00173	0.00109	1.59	0.112
friends__count__wave__1	0.042	0.0142	2.96	0.00322
birth__year	-5.11e-05	0.0017	-0.0301	0.976
family__income	0.00626	0.0068	0.921	0.358
education	-0.00593	0.0169	-0.35	0.726
gender	-0.0277	0.0462	-0.6	0.549

	Estimate	Std..Error	t.value	Pr... t..
northeast	0.0211	0.0718	0.294	0.769
north_central	-0.0873	0.0679	-1.29	0.199
south	-0.0582	0.0593	-0.982	0.327

```r
#Democrats

democrat_ITT_model<-lm(substantive_ideology_scale_wave_5~
                      #treatment assignment variable
                      treat+
                      #pre-treatment ideology score
                      substantive_ideology_scale_wave_1+
                      #% of people followed on Twitter from same party
                      percent_co_party+
                      #% of people in offline networks from same party
                      ideo_homogeneity_offline+
                      #total number of people followed pre-treatment
                      friends_count_wave_1+
                      #demographics
                      birth_year +
                      family_income+
                      education+
                      gender+
                      northeast+
                      north_central+
                      south+
                      #factor variable used to create treatment blocks
                      as.factor(bin_maker),
                      data=democrats)

coefficients<-data.frame(summary(democrat_ITT_model, cluster="bin_maker")$coefficients[2:13,])

library(pander)
panderOptions('digits',3)
panderOptions('table.split.table', 300)
set.caption("Intent-to-Treat Model (Democrats)")
pander(coefficients)
```

Table 7: Intent-to-Treat Model (Democrats)

	Estimate	Std..Error	t.value	Pr…t..
treat	-0.0248	0.0326	-0.76	0.447
substantive_ideology_scale_wave_1	0.834	0.0225	37	3.48e-156
percent_co_party	-0.109	0.0903	-1.21	0.228
ideo_homogeneity_offline	0.000554	0.000734	0.755	0.451
friends_count_wave_1	-0.000722	0.0121	-0.0595	0.953
birth_year	-0.00107	0.00113	-0.951	0.342
family_income	0.00158	0.00502	0.316	0.752
education	0.00212	0.0139	0.153	0.879
gender	0.031	0.0329	0.94	0.348
northeast	-0.0497	0.047	-1.06	0.291
north_central	0.0194	0.0463	0.42	0.675
south	-0.0152	0.0418	-0.364	0.716

5.2 Complier Average Causal Effects

We calculated the Complier Average Causal Effect (CACE) using the two-stage least squares approach developed by Imbens and Rubin (2015), with cluster-robust standard errors to account for our block design (Abadie et al. 2017). In the models reported in the main text of our manuscript, we report results for respondents who were both fully compliant (indicated by answering all weekly compliance checks correctly), and partially compliant (those who answered more than one of the weekly compliance check questions correctly but less than six).

The following assumptions are required to estimate CACE: 1) Ignorability, 2) Monotonicity, 3) Stable Unit Treatment Value, 4) Non-Interference, and, 5) Excludability. The first and second of these assumptions are supported by our research design, and we believe the third and fourth assumptions are warranted because of the extensive steps we took to eliminate respondents from the initial pre-treatment survey who followed each other on Twitter. The excludability assumption—or the assumption that those who did not comply with treatment have the same potential outcomes as those in control—is more problematic. We believe this assumption is warranted for our most basic definition of compliance: whether or not respondents assigned to treatment accepted our invitation to follow the bot. Because we were able to monitor who was following the bot at all times, it is unlikely that anyone who was invited to follow the bot but did not do so was ultimately exposed to its messages—particularly in light of our aforementioned attempts to mitigate causal interference. On the other hand, the excludability assumption is much less strong for our two other compliance measures, which describe whether respondents who followed the bot were able to answer some or all of the questions our surveys asked them about the bot's tweets. This is because it is likely that some of the respondents who accepted our invitation to follow the bot but did not answer any questions correctly about the content of its tweets were nevertheless exposed to some of its messages. Because this assumption is rather strong, we provide multiple estimates of our treatment effects (see above), and encourage readers to focus on our most basic compliance measure (whether or not respondents' accepted our invitation to follow one of the study's two bots)

The code below was used to calculate CACE for the three different levels of compliance described above:

```
#create list of datasets
datasets<-list(democrats, republicans)

#create function to calculate Complier Average Causal effect for
#fully compliant respondents

library(ivpack)
CACE_fc<-function(data){
  #drop cases without outcome response for final survey
  data<-data[complete.cases(data),]

  results<-ivreg(substantive_ideology_scale_wave_5 ~
                      perfect_complier+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker)
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker),
                      data = data)
      #calculate cluster robust standard errors
      data$bin_maker<-as.numeric(data$bin_maker)
      output<-cluster.robust.se(results, data$bin_maker)[2,]
      return(output)}


#create function to calculate Complier Average Causal Effect for
#partially compliant respondents

CACE_hc<-function(data){
  #drop cases without outcome response for final survey
  data<-data[complete.cases(data),]

  results<-ivreg(substantive_ideology_scale_wave_5 ~
                      half_complier+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker)
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker),
                      data = data)
      #calculate cluster robust standard errors
      data$bin_maker<-as.numeric(data$bin_maker)
      output<-cluster.robust.se(results, data$bin_maker)[2,]
      return(output)}


#create function to calculate Complier Average Causal Effect for
#respondents who followed bot

CACE_bf<-function(data){
  #drop cases without outcome response for final survey
  data<-data[complete.cases(data),]

  results<-ivreg(substantive_ideology_scale_wave_5 ~
                      bot_followers+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker)
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      as.factor(bin_maker),
                      data = data)
      #calculate cluster robust standard errors
      data$bin_maker<-as.numeric(data$bin_maker)
      output<-cluster.robust.se(results, data$bin_maker)[2,]
      return(output)}
```

```
#run models
full_compliance_models <- lapply(datasets, function(x) CACE_fc(x))
half_compliance_models <- lapply(datasets, function(x) CACE_hc(x))
bot_follower_models <- lapply(datasets, function(x) CACE_bf(x))

#extract results for republicans
republican_full_compliance_cace<-as.data.frame(t(full_compliance_models[[2]]))
republican_full_compliance_cace$sample<-"republicans_full_compliance"
republican_full_compliance_cace$party<-"republicans"
names(republican_full_compliance_cace)<-c("estimate","se","t","p","sample","party")
republican_half_compliance_cace<-as.data.frame(t(half_compliance_models[[2]]))
republican_half_compliance_cace$sample<-"republicans_half_compliance"
republican_half_compliance_cace$party<-"republicans"
names(republican_half_compliance_cace)<-c("estimate","se","t","p","sample","party")
republican_bot_follower_cace<-as.data.frame(t(bot_follower_models[[2]]))
republican_bot_follower_cace$sample<-"republicans_bot_follower"
republican_bot_follower_cace$party<-"republicans"
names(republican_bot_follower_cace)<-c("estimate","se","t","p","sample","party")

#extract results for democrats
democrat_full_compliance_cace<-data.frame(t(full_compliance_models[[1]]))
democrat_full_compliance_cace$sample<-"democrats_full_compliance"
democrat_full_compliance_cace$party<-"democrats"
names(democrat_full_compliance_cace)<-c("estimate","se","t","p","sample","party")
democrat_half_compliance_cace<-as.data.frame(t(half_compliance_models[[1]]))
democrat_half_compliance_cace$sample<-"democrats_half_compliance"
democrat_half_compliance_cace$party<-"democrats"
names(democrat_half_compliance_cace)<-c("estimate","se","t","p","sample","party")
democrat_bot_follower_cace<-as.data.frame(t(bot_follower_models[[1]]))
democrat_bot_follower_cace$sample<-"democrats_bot_follower"
democrat_bot_follower_cace$party<-"democrats"
names(democrat_bot_follower_cace)<-c("estimate","se","t","p","sample","party")
```

The following code was used to produce Figure 1 in the main text of our article.

```
#create another dataset that combines ITT and CACE results for plotting

republican_itt<-
  data.frame(t(summary(republican_ITT_model, cluster="bin_maker")$coefficients[2:2,]))
names(republican_itt)<-c("estimate","se","t","p")
republican_itt$sample<-"republicans_itt"
republican_itt$party<-"republicans"

democrat_itt<-
  data.frame(t(summary(democrat_ITT_model, cluster="bin_maker")$coefficients[2:2,]))
names(democrat_itt)<-c("estimate","se","t","p")
democrat_itt$sample<-"democrats_itt"
democrat_itt$party<-"democrats"


republican_plot<-rbind(republican_full_compliance_cace,
                       republican_half_compliance_cace,
                       republican_bot_follower_cace,
                       republican_itt)

democrat_plot<-rbind(democrat_full_compliance_cace,
                     democrat_half_compliance_cace,
                     democrat_bot_follower_cace,
                     democrat_itt)



republican_plot$sample<-factor(republican_plot$sample,
                               levels=c("republicans_full_compliance",
                                 "republicans_half_compliance",
                                 "republicans_bot_follower",
                                 "republicans_itt"),
                               labels=c("Fully Compliant Respondents",
                                  "Partially Compliant Respondents",
                                  "Minimally Compliant Respondents",
                                  "Respondents Assigned to Treatment"))

democrat_plot$sample<-factor(democrat_plot$sample,
                               levels=c("democrats_full_compliance",
                                 "democrats_half_compliance",
                                 "democrats_bot_follower",
                                 "democrats_itt"),
                               labels=c("Fully Compliant Respondents",
                                  "Partially Compliant Respondents",
                                  "Minimally Compliant Respondents",
                                  "Respondents Assigned to Treatment"))


#create standard error bars
interval1 <- -qnorm((1-0.9)/2)  # 90% multiplier
interval2 <- -qnorm((1-0.95)/2)  # 95% multiplier

#create plot
library(ggplot2)
figure_1_dems<-ggplot(democrat_plot)+
  geom_hline(yintercept = 0, colour = gray(1/2), lty = 2)+
  geom_point(aes(x=sample, y=estimate),
             position = position_dodge(width = 1/2),
             size=2, colour="blue")+
  geom_linerange(aes(x = sample, ymin = estimate - se*interval1,
                     ymax = estimate + se*interval1),
                 lwd = 1, position = position_dodge(width = 1/2),
                 colour="blue")+
  geom_linerange(aes(x = sample, y = estimate, ymin = estimate - se*interval2,
                     ymax = estimate + se*interval2),
                 lwd = .5, position = position_dodge(width = 1/2),
                 colour="blue")+
  theme(axis.text=element_text(size=9, face="bold",colour="black"),
        plot.title = element_text(face="bold", size=16, hjust = 0.5,vjust=3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background=element_blank(),
        axis.title=element_text(size=9, colour="black"),
        legend.position="none",
        legend.key = element_blank(),
        legend.title=element_blank())+
  ylim(c(-1,1))+
  labs(x="",y="")+
  coord_flip()+
  ggtitle("Democrats")


figure_1_reps<-ggplot(republican_plot)+
  geom_hline(yintercept = 0, colour = gray(1/2), lty = 2)+
  geom_point(aes(x=sample, y=estimate),
             position = position_dodge(width = 1/2),
             size=2, colour="red")+
  geom_linerange(aes(x = sample, ymin = estimate - se*interval1,
                     ymax = estimate + se*interval1),
                 lwd = 1, position = position_dodge(width = 1/2),
                 colour="red")+
  geom_linerange(aes(x = sample, y = estimate, ymin = estimate - se*interval2,
                     ymax = estimate + se*interval2),
                 lwd = .5, position = position_dodge(width = 1/2),
                 colour="red")+
  theme(axis.text=element_text(size=9, face="bold",colour="black"),
        plot.title = element_text(face="bold", size=16, hjust = 0.5, vjust=3),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background=element_blank(),
        axis.title=element_text(size=9, colour="black"),
        legend.position="none",
        legend.key = element_blank(),
        legend.title=element_blank())+
  labs(x="",y="")+
  ylim(c(-1,1))+
  coord_flip()+
  ggtitle("Republicans")

ggsave(figure_1_dems, file="Figure 1 dems.png", width=7, height=4, dpi=1000)
ggsave(figure_1_reps, file="Figure 1 reps.png", width=7, height=4, dpi=1000)
```

5.3 Intent-to-Treat Effects without Covariates

In order to calculate the most basic, or "pure," estimate of our causal effect, we also calculated Intent-to-Treat effects without covariates using fisher's exact score to account for the variables we used to define treatment blocks. According to this analysis, Republicans assigned to treatment increased .10 points on our liberal-conservative scale (p<.01), where higher values indicate increased conservatism. Democrats, for their part, decreased .03 points (p<.84)

```
#create function to calculate fisher's exact score
#with grouping variable to account for blocking
```

```r
fisher.exact <- function(Yobs,Wobs,Gobs){
  #Yobs is the observed outcome
  #Wobs is the observed randomziation (T/C)
  #Gobs is the group assignment
  tmp <- which(is.na(Yobs) | is.na(Wobs) | is.na(Gobs))
  Yobs <- Yobs[-tmp]
  Wobs <- Wobs[-tmp]
  Gobs <- Gobs[-tmp]
  J <- 1000
  Yfull <- matrix(c(Yobs,Yobs),nrow=length(Yobs),ncol=2, byrow=FALSE)
  tmp <- NULL
  for(i in 1:J){
    Wtmp <- rep(0,length(Yobs))
    track_means_treat <- track_means_control <- bin_size <- NULL
    for(g in unique(Gobs)){
      Wtmp[Gobs==g] <- c(rep(0,sum(1-Wobs[Gobs==g])),
                         rep(1,sum(Wobs[Gobs==g])))[sample(1:sum(Gobs==g))]
      track_means_treat <- c(track_means_treat,mean(Yfull[Wtmp==1 & Gobs==g,2]))
      track_means_control <- c(track_means_control,mean(Yfull[Wtmp==0 & Gobs==g,1]))
      bin_size <- c(bin_size,sum(Gobs==g))
    }
    tmp <- c(tmp,weighted.mean(track_means_treat,bin_size)-
               weighted.mean(track_means_control,bin_size))
  }
  track_means_treat <- track_means_control <- bin_size <- NULL
  for(g in unique(Gobs)){
    track_means_treat <- c(track_means_treat,mean(Yfull[Wobs==1 & Gobs==g,2]))
    track_means_control <- c(track_means_control,mean(Yfull[Wobs==0 & Gobs==g,1]))
    bin_size <- c(bin_size,sum(Gobs==g))
  }
  tobs = weighted.mean(track_means_treat,bin_size)-
    weighted.mean(track_means_control,bin_size)
  list(null = tmp,tobs = tobs)
}

#Subset Republicans and Democrats

republicans <- final_data[final_data$party_id_wave_1==2,]
democrats <- final_data[final_data$party_id_wave_1==1,]

#Republicans

fisher_exact<-fisher.exact(republicans$substantive_ideology_change,
                           republicans$treat,
                           republicans$bin_maker)

#ATE
republican_itt<-fisher_exact$tobs

#p-Value
republican_itt_p_value<-mean(fisher_exact$null>fisher_exact$tobs)
```

```
#Democrats

fisher_exact<-fisher.exact(democrats$substantive_ideology_change,
                           democrats$treat,
                           democrats$bin_maker)

#ATE
democrat_itt<-fisher_exact$tobs

#p-Value
democrat_itt_p_value<-mean(fisher_exact$null>fisher_exact$tobs)
```

5.4 Interpretation of Effect Size

To further evaluate the effect sizes reported in the main text of our article, we offer an approximate comparison of our findings to historical shifts in the liberal/conservative scale described above, which has been administered sixteen times to a representative sample of American adults between 1994 and 2014 (Dimock and Carroll 2014). Our results are not directly comparable to these previous surveys for several reasons. Previous studies asked respondents ten questions about a social policy issue, each of which had a liberal and conservative-leaning statements, and the respondent was asked to place themselves on a continuum between the two. In contrast, we randomly selected five of these liberal statements and five conservative statements and then asked respondents to agree or disagree with them on a seven-point scale. These divergent scales result in different variance structures which makes a direct comparison of the effects impossible.

The table below presents data from Dimock and Carroll (2014) that aggregates the ideological consistency score into six categories that describe the percentage of the American population that is liberal or conservative. These results were created by combining binary responses into a ten-point liberal (-10) to conservative (+10) scale and then examining the distribution of responses within six quantiles.

% Who are	1994	1999	2004	2011	2014
Consistently Conservative	7	4	3	7	9
Mostly Conservative	23	16	15	19	18
Mixed	49	49	49	49	42
Mostly Liberal	18	25	25	23	22
Consistently Liberal	3	6	8	8	12
Mean Score (-10 to 10)	.6	-.6	-.9	-.3	-.6

Table S1: Distribution of Liberal/Conservative Index Over Time (Dimock and Carroll 2014)

To create an approximate comparison of the size of the conservative backfire effect for fully compliant respondents that we report in our main paper (unstandardized beta=.60) to these data, one can convert the former into a twenty-point scale (20*.60)/7=1.71. To the extent these metrics can be compared in light of the aforementioned scaling issues, this would indicate a shift in attitudes that is substantially larger than that which occurred between 1994 and 2014.

5.5 Using Recursive Partioning to Detect Causal Heterogeneity

We conducted additional analyses to detect possible causal heterogeneity using Athey and Imben's (2016) machine learning approach that employs recursive partitioning. Below we report the result of the LASSO model with change in liberal/conservative ideology scale between the pre and post-treatment waves as the outcome.

```
# Subset variables
  tab = apply(twitter_data[, c(6:20, 59:104)], 2, table)

  for (i in 1:length(tab)){
    print(paste(i, names(tab)[i]))
    print(tab[[i]])
  }

  names(tab)[which(sapply(tab, function(x) length(x) > 10))]
  names(tab)[which(sapply(tab, function(x) length(x) == 1))]
  names(tab)[which(sapply(tab, function(x) length(x) == 0))]

  var_con <- c("birth_year", "followers_count_wave_1", "statuses_count_wave_1",
               "friends_count_wave_1", "family_income")

  var_nom <- c(names(tab)[-c(which(sapply(tab, function(x) length(x) > 10)),
                    which(sapply(tab, function(x) length(x) <= 1)))],
               "state", "religion", "protestant_church", "naics_industry_code")

  # Handling missing values
  apply(twitter_data[, var_con], 2, function(x) sum(is.na(x)))

  count_miss <- apply(twitter_data[, var_nom], 2, function(x) sum(is.na(x)))
  count_miss[which(count_miss > 0)]

  var_nom <- var_nom[which(!var_nom %in% c("protestant_church"))]

  x_con <- twitter_data[, var_con]
  x_nom <- data.frame(apply(twitter_data[, var_nom], 2,
                      function(x) factor(as.character(x),
                    levels = names(table(x, useNA = "ifany")))))
  x <- as.data.frame(model.matrix(~.+0, data = as.data.frame(cbind(x_nom, x_con))))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
## -2.10000 -0.30000  0.00000 -0.00834  0.20000  2.20000      141
```

```
## [1] 956
```

```
## Loading required package: Matrix
```

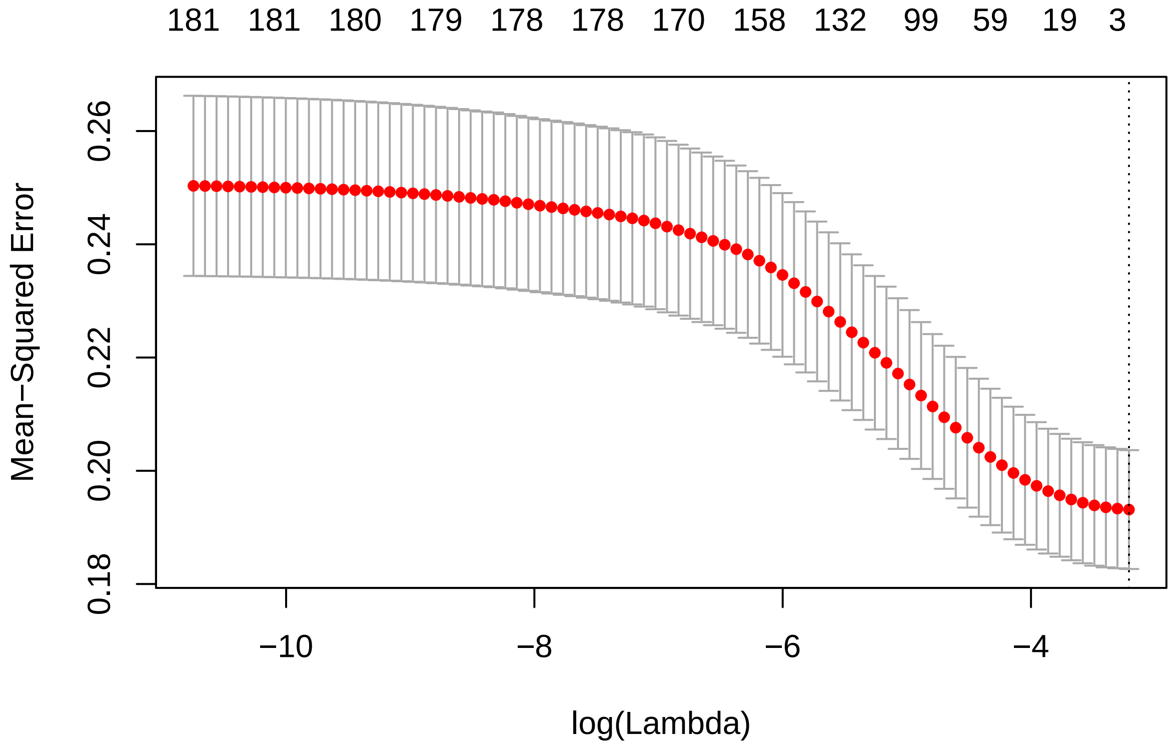
```
## Loading required package: foreach
```

```
## Loaded glmnet 2.0-13
```

```
coef(cvfit.lasso, s = "lambda.min")
```

```
## 190 x 1 sparse Matrix of class "dgCMatrix"
##                                                          1
## (Intercept)                                   -0.008338729
## newsint_wave_11                                .
## newsint_wave_12                                .
## newsint_wave_13                                .
## newsint_wave_14                                .
## newsint_wave_17                                .
## news_source_newspaper_hard_copy_wave_12        .
## news_source_newspapers_website_wave_12          .
## news_source_news_website_not_newspaper_wave_12  .
## news_source_news_app_mobile_device_wave_12      .
## news_source_email_newsletters_RSS__wave_12      .
## news_source_social.network_websites_wave_12     .
## news_source_blogs_not_major_media_wave_12       .
## news_source_television_wave_12                  .
## news_source_radio_wave_12                       .
## news_source_magazines_wave_12                   .
## news_source_podcasts_wave_12                    .
## news_source_other_wave_12                       .
## news_source_none_of_the_above_wave_12           .
## news_source_dont_know_wave_12                   .
## gender2                                         .
## race2                                           .
## race3                                           .
## race4                                           .
## race5                                           .
## race6                                           .
## race7                                           .
## race8                                           .
## hispanic2                                       .
## multrace_white2                                 .
## multrace_black2                                 .
## multrace_hispanic2                              .
## multrace_asian2                                 .
## multrace_native_american2                       .
## multrace_middle_eastern2                        .
## multrace_dont_know2                             .
## education2                                      .
## education3                                      .
## education4                                      .
## education5                                      .
## education6                                      .
## marital_status2                                 .
## marital_status3                                 .
## marital_status4                                 .
## marital_status5                                 .
## marital_status6                                 .
## has_children_under_18 2                         .
## speaks_panish 2                                 .
## speaks_panish 3                                 .
## speaks_panish 4                                 .
## employed2                                       .
## employed3                                       .
## employed4                                       .
## employed5                                       .
## employed6                                       .
## employed7                                       .
## employed8                                       .
## employed9                                       .
## employment_otherSelf Employed                   .
## industrynaicsotherAutomotive Manufacturing      .
## industrynaicsotherinformation techology         .
## industrynaicsotherPublic Affairs                .
## industrynaicsotherRetail                        .
## industrynaicsotherWriting                       .
## voter_registration_status2                      .
## voter_registration_status3                      .
## presidential_vote_2016 2                        .
## presidential_vote_2016 3                        .
## presidential_vote_2016 4                        .
## presidential_vote_2016 5                        .
## presidential_vote_2016 6                        .
## presidential_vote_2016 7                        .
## ideology2                                       .
## ideology3                                       .
## ideology4                                       .
## ideology5                                       .
## ideology6                                       .
## born_again2                                     .
## important_religion2                             .
## important_religion3                             .
## important_religion4                             .
## church_attendance2                              .
## church_attendance3                              .
## church_attendance4                              .
## church_attendance5                              .
## church_attendance6                              .
## church_attendance7                              .
## frequency_of_prayer2                            .
## frequency_of_prayer3                            .
## frequency_of_prayer4                            .
## frequency_of_prayer5                            .
## frequency_of_prayer6                            .
## frequency_of_prayer7                            .
## frequency_of_prayer8                            .
## religpew_t__NA__                                .
## religpew_tBaptist                               .
## religpew_tChristian                             .
## religpew_tEpiscopalian                          .
## religpew_tMethodist                             .
## religpew_tQuaker                                .
## religpew_protestant_tNONE                       .
## republican_wave_11                              .
## state 4                                         .
## state 5                                         .
## state 6                                         .
## state 8                                         .
## state 9                                         .
## state10                                         .
## state11                                         .
## state12                                         .
## state13                                         .
## state15                                         .
## state16                                         .
## state17                                         .
## state18                                         .
## state19                                         .
## state20                                         .
## state21                                         .
## state22                                         .
## state23                                         .
## state24                                         .
## state25                                         .
## state26                                         .
## state27                                         .
## state28                                         .
## state29                                         .
## state30                                         .
## state31                                         .
## state32                                         .
## state33                                         .
## state34                                         .
## state35                                         .
## state36                                         .
## state37                                         .
## state38                                         .
## state39                                         .
## state40                                         .
## state41                                         .
## state42                                         .
## state44                                         .
## state45                                         .
## state46                                         .
## state47                                         .
## state48                                         .
## state49                                         .
## state50                                         .
## state51                                         .
## state53                                         .
## state54                                         .
## state55                                         .
## state56                                         .
## religion 2                                      .
## religion 3                                      .
## religion 4                                      .
## religion 5                                      .
## religion 6                                      .
## religion 7                                      .
## religion 8                                      .
## religion 9                                      .
## religion10                                      .
## religion11                                      .
## religion12                                      .
## naics_industry_code 2                           .
## naics_industry_code 3                           .
## naics_industry_code 4                           .
## naics_industry_code 5                           .
## naics_industry_code 6                           .
## naics_industry_code 7                           .
## naics_industry_code 8                           .
## naics_industry_code 9                                .
## naics_industry_code10                                .
## naics_industry_code11                                .
## naics_industry_code12                                .
## naics_industry_code13                                .
## naics_industry_code14                                .
## naics_industry_code15                                .
## naics_industry_code16                                .
## naics_industry_code17                                .
## naics_industry_code18                                .
## naics_industry_code19                                .
## naics_industry_code20                                .
## naics_industry_code21                                .
## naics_industry_code22                                .
## naics_industry_code23                                .
## naics_industry_code99                                .
## birth_year                                           .
## followers_count_wave_1                               .
## statuses_count_wave_1                                .
## friends_count_wave_1                                 .
## family_income                                        .
```

```
plot(cvfit.lasso)
```



6 Additional Robustness Checks

6.1 Attrition Bias

The table below indicates there is no evidence of attrition bias by treatment condition.

Condition	Pre-Treatment Survey	Post-Treatment Survey
Control	495	413
Treatment	744	656

The table below describes the demographic characteristics of respondents in the pre and post-treatment surveys as well as the three compliance check surveys that were administered during the one month study period

Variable	Pre-Treatment Survey	Post-Treatment Survey
% Republican	43%	42.1%
% Female	51.9%	51.8%
Age (mean)	50.5	51.6
% Northeast	20.3%	21.0%
% South	39.5%	39.2%
% North Central	20.0%	20.4%
% West	21.4%	21.0%

In order to further examine attrition bias created by any of the covariates in our model, we employed the pooling test created by Becketti, Gould, Lillard, and Welch (1988). First, we regressed our outcome variable on the control variables employed in our main analyses alongside a binary indicator variable that describes whether the respondent completed the post-treatment survey. Next, we ran an identical model with interaction terms between the attrition indicator and each of the other variables in the model. An F-test indicates there is no significant difference between the models for Democrats (F=0.85, p<.60) or Republicans (F=1.44, p<.13), suggesting there is no evidence of attrition bias according to these covariates. The code below as used to perform this analysis:

```
#create missing value indicator for post-treatment survey
final_data$wave_5_missing<-0
final_data$wave_5_missing[is.na(final_data$endtime_wave_5)]<-1
table(final_data$wave_5_missing)

#analyze Republican experiment
republicans<-final_data[final_data$party_id_wave_1==2,]

reduced_model<-lm(substantive_ideology_scale_wave_1~
                 wave_5_missing+
                 percent_co_party+
                 political_wave_1+
                 freq_twitter_wave_1+
                 friends_count_wave_1+
                 strong_partisan+
                 birth_year +
                 family_income+
                 education+
                 gender+
                 ideo_homogeneity_offline+
                 northeast+
                 north_central+
                 south,
                 data=republicans)
```

```
full_model<-lm(substantive_ideology_scale_wave_1~
               wave_5_missing+
               percent_co_party+
               political_wave_1+
               freq_twitter_wave_1+
               friends_count_wave_1+
               strong_partisan+
               birth_year +
               family_income+
               education+
               gender+
               ideo_homogeneity_offline+
               northeast+
               north_central+
               south+
               wave_5_missing*percent_co_party+
               wave_5_missing*political_wave_1+
               wave_5_missing*freq_twitter_wave_1+
               wave_5_missing*friends_count_wave_1+
               wave_5_missing*strong_partisan+
               wave_5_missing*birth_year +
               wave_5_missing*family_income+
               wave_5_missing*education+
               wave_5_missing*gender+
               wave_5_missing*ideo_homogeneity_offline+
               wave_5_missing*northeast+
               wave_5_missing*north_central+
               wave_5_missing*south,
               data=republicans)

anova(reduced_model, full_model)


#analyze Democrat Experiment
democrats<-final_data[final_data$party_id_wave_1==1,]

reduced_model<-lm(substantive_ideology_scale_wave_1~
                 wave_5_missing+
                 percent_co_party+
                 political_wave_1+
                 freq_twitter_wave_1+
                 friends_count_wave_1+
                 strong_partisan+
                 birth_year +
                 family_income+
                 education+
                 gender+
                 ideo_homogeneity_offline+
                 northeast+
                 north_central+
                 south,
                 data=democrats)
```

```
full_model<-lm(substantive_ideology_scale_wave_1~
               wave_5_missing+
               percent_co_party+
               political_wave_1+
               freq_twitter_wave_1+
               friends_count_wave_1+
               strong_partisan+
               birth_year +
               family_income+
               education+
               gender+
               ideo_homogeneity_offline+
               northeast+
               north_central+
               south+
               wave_5_missing*percent_co_party+
               wave_5_missing*political_wave_1+
               wave_5_missing*freq_twitter_wave_1+
               wave_5_missing*friends_count_wave_1+
               wave_5_missing*strong_partisan+
               wave_5_missing*birth_year +
               wave_5_missing*family_income+
               wave_5_missing*education+
               wave_5_missing*gender+
               wave_5_missing*ideo_homogeneity_offline+
               wave_5_missing*northeast+
               wave_5_missing*north_central+
               wave_5_missing*south,
               data=democrats)

anova(reduced_model, full_model)
```

6.2 Experiment Effects

As with all field experiments, it is possible that respondents in our study shifted their behavior in response to being part of a study. Though such experiment effects typically lead respondents to change their behavior in line with what they perceive to be the expectations of the researchers, it is possible that Republicans in our study had the opposite reaction. That is, Republicans may have expressed more conservative views because they thought the purpose of the study was to make them more liberal, and responded by expressing more conservative views.

As we mentioned in our description of the treatment delivery above, we took several steps to mitigate the likelihood of experiment effects. First, we employed open-ended questions in two pilot studies that asked people to guess the purpose of our study. Because several respondents in our pilot studies guessed correctly, we made the following shifts to our research design. First, we employed an ostensibly unrelated survey design in which the treatment delivery occurred one week after the pre-treatment survey (Broockman and Kalla 2016). Second, respondents were only shown pictures of landscapes for the first few days of the study, and tweets from those with opposing ideologies were gradually inserted into their Twitter feed thereafter. Third, we revised our informed consent dialogue so that it did not disclose that we were academics, but rather "a research sponsor" (though the IRB contact information for the first author's university remained in the last paragraph of the informed consent dialogue).

Thus, in order to respond in an expressive manner to our study, respondents would have to a) connect the invitation to follow one of our Twitter bots to the pre-treatment and post-treatment surveys; b) remember how strongly they agreed or disagreed with statements on the pre-treatment survey in order to increase them accordingly (since we estimate a difference-in-difference effect); and, c) read the entire informed consent dialogue carefully—at least insofar as the expressive effects described above were a reaction to us as researchers, and not the treatment itself.

Though we cannot rule out the possibility of experiment effects entirely, we believe they are unlikely for the following reasons. First, according to the YouGov Director of Scientific Research, most of the people in the regular YouGov panel take multiple surveys each week—many of which ask them questions about politics. Presumably, connecting our pre-treatment survey to the invitation to follow the Twitter bot would be rather difficult. Second, if people wished to respond in an expressive manner, remembering how they responded to the initial survey would be difficult after one month—particularly for those who regularly take surveys about political issues online. Presumably, many of those who wish to respond expressively would choose the highest category of response—strongly agreeing with conservative leaning questions and strongly disagreeing with liberal questions after providing answers of different strength in the first wave—in order to demonstrate their displeasure. None of the respondents in our study provided this type of response. Finally, the size of the backfire effect we observed increases with level of compliance—whereas one would expect fully compliant respondents to be no more likely to respond in an expressive manner than partially compliant respondents.

6.3 Outliers

We examined the robustness of our findings to outliers by using Cook's Distance to identify 21 cases which were four times the mean value of Cook's distance for all observations. These findings show the effect reported in the remain paper is robust to the exclusion of these outliers.

```
first.stage.1 <- lm(substantive_ideology_scale_wave_5 ~
                      treat +
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      birth_year +
                      family_income +
                      education +
                      gender +
                      ideo_homogeneity_offline +
                      northeast +
                      north_central +
                      south,
                    data=republicans,
                    na.action=na.exclude)
summary(first.stage.1)

republicans$instrumented.perfcomp <- fitted(first.stage.1)

second.stage.1 <- lm(substantive_ideology_scale_wave_5 ~
                       instrumented.perfcomp +
                       substantive_ideology_scale_wave_1+
                       percent_co_party +
                       political_wave_1 +
                       freq_twitter_wave_1 +
                       friends_count_wave_1+
                       strong_partisan +
                       birth_year +
                       family_income +
                       education +
                       gender +
                       ideo_homogeneity_offline +
                       northeast +
                       north_central +
                       south,
                     data=republicans,
                     na.action=na.exclude)

summary(second.stage.1)

cooksd <- cooks.distance(second.stage.1)
plot(cooksd, pch="*", cex=2, main="Influential Obs by Cooks distance")
influential <- as.numeric(na.omit(names(cooksd)[(cooksd > 4*mean(cooksd, na.rm=T))]))
length(influential)
# identifies 16 cases above accepted val

republicans$rownumber <- as.numeric(rownames(republicans))
rep_rmalloutlier <- republicans[ ! republicans$rownumber %in% influential,]

#subset variables for models

rep_rmalloutlier<-rep_rmalloutlier[,c("substantive_ideology_scale_wave_5",
                      "perfect_complier",
                      "treat",
                      "substantive_ideology_scale_wave_1",
                      "bin_maker",
                      "percent_co_party",
                      "friends_count_wave_1",
                      "birth_year",
                      "family_income",
                      "education",
                      "gender",
                      "ideo_homogeneity_offline",
                      "northeast",
                      "north_central",
                      "south")]



rep_rmalloutlier<-rep_rmalloutlier[complete.cases(rep_rmalloutlier),]


library(ivpack)
outliers_removed<-ivreg(substantive_ideology_scale_wave_5 ~
                      perfect_complier+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south,
                      data = rep_rmalloutlier)
      #calculate cluster robust standard errors
      rep_rmalloutlier$bin_maker<-as.numeric(rep_rmalloutlier$bin_maker)
      output<-cluster.robust.se(outliers_removed, rep_rmalloutlier$bin_maker)[2,]
      output
```

6.4 Compound Treatment

Though our various measures of compliance with treatment enable us to make a more accurate estimate of the effect of exposure to Twitter accounts with opposing ideological views, our use of financial incentives to encourage people to pay attention to Twitter might have also increased their exposure to Twitter more broadly, and particularly Twitter messages about current events. This might create spillover effects which compound our treatment (i.e. the backfire effect might not result from exposure to Twitter accounts with opposing ideological views, but because of exposure to more information about current events in general). Though we are not aware of any studies that indicate such exposure can shape opinions on the types of policy issues we measure in this study, there are a variety of other studies which indicate such exposure has a negative effect on political participation (Eliasoph 1998). To examine whether spillover effects occurred, we examined whether compliance with our treatment increased the frequency with which respondents report using Twitter. As the code below shows, we detected no significant evidence of spillover effects using this strategy (Full Compliers, CACE: t=1.28, p<.19, Partial Compliers: CACE: t=1.29, p<.20)

```
twitter_data$twitter_use_change<-twitter_data$how_often_visit_twitter_wave_1-
    twitter_data$how_often_visit_twitter_wave_5

library(ivpack)
summary(ivreg(twitter_use_change~
                perfect_complier
              |
                treat,
              data=twitter_data))

summary(ivreg(twitter_use_change~
                half_complier
              |
                treat,
              data=twitter_data))
```

6.5 Post-estimation Weighting by Age

Readers may recall that the demographic characteristics of respondents in our sample track national averages quite well. Because our sample is slighty older than the general U.S. population (median age of our respondents: 50.48, median age of U.S. over-18 population: 46.37), however, it is possible that backfire effects among older Republican respondents drive our findings. To account for this, we ran CACE models with weights for age using data from the American Community Survey. These analyses produced nearly identical results.

```
census_age<-read.csv("~/Desktop/census age data.csv", stringsAsFactors = FALSE)
weights<-census_age[,c("age_group","both_sexes_percent_of_pop")]
names(weights)<-c("age","percent_pop")

final_data$age_percent<-NA
for (i in 1:nrow(final_data)){
final_data$age_percent[i]<-
  nrow(final_data[final_data$birth_year==final_data$birth_year[i],])/
  nrow(final_data)
print(i)
}

library(dplyr)
final_data$age<-2017-final_data$birth_year
weights$age<-gsub(" years","", weights$age)
weights$age<-as.numeric(weights$age)
for_weight<-left_join(final_data, weights)

for_weight$weight<-for_weight$age_percent/for_weight$percent_pop

#subset republicans

republicans<-for_weight[for_weight$party_id_wave_1==2,]
republicans<-republicans[complete.cases(republicans),]

weighted_model<-ivreg(substantive_ideology_scale_wave_5 ~
                      perfect_complier+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south,
                      data = republicans,
                      weights=weight)

cluster.robust.se(weighted_model, as.numeric(republicans$bin_maker))

democrats<-for_weight[for_weight$party_id_wave_1==1,]
democrats<-democrats[complete.cases(democrats),]

weighted_model_dems<-ivreg(substantive_ideology_scale_wave_5 ~
                      perfect_complier+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year +
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south
                      |
                      treat+
                      substantive_ideology_scale_wave_1+
                      as.factor(bin_maker)+
                      percent_co_party+
                      friends_count_wave_1+
                      birth_year+
                      family_income+
                      education+
                      gender+
                      ideo_homogeneity_offline+
                      northeast+
                      north_central+
                      south+
                      weight,
                      data = democrats,
                      weights=weight)

cluster.robust.se(weighted_model_dems, as.numeric(democrats$bin_maker))
```

References

Abadie, Alberto, Susan Athey, Guido Imbens, and Jeffrey Wooldridge. 2017. "When Should You Adjust Standard Errors for Clustering?" *arXiv:1710.02926*.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113 (27). National Acad Sciences: 7353–60.

Bafumi, Joseph, and Michael C Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress." *American Political Science Review* 104 (3). Cambridge University Press: 519–42.

Barberá, Pablo. 2015. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23 (1): 76–91. doi:10.1093/pan/mpu011.

Broockman, David, and Joshua Kalla. 2016. "Durably reducing transphobia: A field experiment on door-to-door canvassing." *Science (New York, N.Y.)* 352 (6282). American Association for the Advancement of Science: 220–4. doi:10.1126/science.aad9713.

Dimock, Michael, and Doherty Carroll. 2014. "Political Polarization in the American Public: How Increasing Ideological Uniformity and Partisan Antipathy Affect Politics, Compromise, and Everyday Life." *Pew Research Center Report*.

Eliasoph, Nina. 1998. *Avoiding Politics: How Americans Produce Apathy in Everyday Life*. Cambridge University Press.

Fiorina, Morris P, Samuel J Abrams, and Jeremy Pope. 2006. *Culture War?: The Myth of a Polarized America*. Longman Publishing Group.

Gentzkow, Matthew, and Jesse M Shapiro. 2011. "Ideological Segregation Online and Offline." *The Quarterly Journal of Economics* 126 (4). MIT Press: 1799–1839.

Grönlund, Kimmo, Kaisa Herne, and Maija Setälä. 2015. "Does Enclave Deliberation Polarize Opinions?" *Political Behavior* 37 (4). Springer: 995–1020.

Gupta, Pankaj, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. "Wtf: The Who to Follow Service at Twitter." In *Proceedings of the 22nd International Conference on World Wide Web*, 505–14. ACM.

Hill, Seth J, and Chris Tausanovitch. 2015. "A Disconnect in Representation? Comparison of Trends in Congressional and Public Polarization." *The Journal of Politics* 77 (4). University of Chicago Press Chicago, IL: 1058–75.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Jessee, Stephen A. 2012. *Ideology and Spatial Voting in American Elections*. Cambridge University Press.

Lazer, David, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176). American Association for the Advancement of Science: 1203–5.

Luskin, Robert C, James S Fishkin, and Kyu S Hahn. 2007. "Deliberation and Net Attitude Change." In *ECPR General Conference, Pisa, Italy*, 6–8.

Tausanovitch, Chris, and Christopher Warshaw. 2013. "Measuring Constituent Policy Preferences in Congress, State Legislatures, and Cities." *The Journal of Politics* 75 (2). Cambridge University Press New York, USA: 330–42.