

Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?



Pablo Barberá¹, John T. Jost^{1,2,3}, Jonathan Nagler³,
Joshua A. Tucker³, and Richard Bonneau⁴

¹Center for Data Science, ²Department of Psychology, ³Department of Politics, and ⁴Center for Genomics and Systems Biology, New York University

Abstract

We estimated ideological preferences of 3.8 million Twitter users and, using a data set of nearly 150 million tweets concerning 12 political and nonpolitical issues, explored whether online communication resembles an “echo chamber” (as a result of selective exposure and ideological segregation) or a “national conversation.” We observed that information was exchanged primarily among individuals with similar ideological preferences in the case of political issues (e.g., 2012 presidential election, 2013 government shutdown) but not many other current events (e.g., 2013 Boston Marathon bombing, 2014 Super Bowl). Discussion of the Newtown shootings in 2012 reflected a dynamic process, beginning as a national conversation before transforming into a polarized exchange. With respect to both political and nonpolitical issues, liberals were more likely than conservatives to engage in cross-ideological dissemination; this is an important asymmetry with respect to the structure of communication that is consistent with psychological theory and research bearing on ideological differences in epistemic, existential, and relational motivation. Overall, we conclude that previous work may have overestimated the degree of ideological segregation in social-media usage.

Keywords

political ideology, polarization, social media, open data, open materials

Received 3/26/15; Revision accepted 6/15/15

You can observe a lot by watching.

—Yogi Berra (2008)

The spread of Internet usage has expanded the quantity and variety of political information to which citizens have access and has created unprecedented opportunities for communicating with peers about current events. Research on the effects of group diversity on decision-making quality (Hong & Page, 2004) has sometimes been taken to suggest that technological transformations should contribute to a more robust and pluralistic form of public debate (Mutz, 2006). However, to the extent that individuals expose themselves to information that simply reinforces their existing views (e.g., Ditto & Lopez, 1992; Garrett, 2009a, 2009b; Sears & Freedman, 1967), greater access to information may foster selective exposure to ideologically congenial content, resulting in an “echo

chamber” environment that could facilitate social extremism and political polarization (Adamic & Glance, 2005; Iyengar & Hahn, 2009; Prior, 2007). By the same token, perceptions of political polarization may be exaggerated for partisan and other reasons (Van Boven, Judd, & Sherman, 2012; Westfall, Van Boven, Chambers, & Judd, 2015).

Previous investigations of selective exposure to media have relied either on self-reported survey responses (Gentzkow & Shapiro, 2011; Pfau, Houston, & Semmler, 2007), which are subject to measurement error and social-desirability bias, or on behavior exhibited in laboratory experiments (Garrett, 2009b; Iyengar & Hahn, 2009; Prior,

Corresponding Author:

Pablo Barberá, Center for Data Science, New York University, 726 Broadway, 7th Floor, New York, NY 10003
E-mail: pablo.barbera@nyu.edu

2007), which may be low in external validity. An ideal setting for the study of mass-level polarization would be one that allows researchers to unobtrusively observe behavioral choices in the environment in which they naturally occur (Webb, Campbell, Schwartz, & Sechrest, 1966) and to infer the characteristics of individuals who engage in such behavior (Back et al., 2010). The use of social-media data satisfies the first condition, but, until recently, estimates of individual-level characteristics of social-media users have been calculated only for small, self-selected samples (Conover et al., 2011; Kosinski, Stillwell, & Graepel, 2013) or have focused on variables that are of little direct relevance to the study of ideological polarization (Golder & Macy, 2011).

In the research reported here, we leveraged a method that generates valid estimates of the ideological positions of extremely large numbers of social-media users, building on techniques developed by Barberá (2015). We exploited an underutilized dimension of social-media data sets, namely, the structures of observable social networks. More specifically, we built on the notion that decisions about whom to “follow” on platforms such as Twitter convey information about individual users’ political preferences. This follows from the notion that individuals tend to interact with others who are similar rather than dissimilar to themselves (McPherson, Smith-Lovin, & Cook, 2001). Our approach built on the logic of latent space models (Hoff, Raftery, & Handcock, 2002) and spatial following models (Barberá, 2015), but we relied on a mathematical approximation using correspondence analysis (Greenacre, 1993) to ensure that our method was computationally tractable even with large-scale networks involving millions of nodes. Furthermore, because our method did not rely on text-processing techniques, it was not subject to methodological limitations that are associated with textual or content analysis.

The extent to which citizens exhibit patterns of ideological polarization in online exchanges remains an open debate. Some researchers have reported high levels of clustering along party lines (Conover, Gonçalves, Flammini, & Menczer, 2012), characterizing social-media platforms as echo chambers, whereas others have found that ideological segregation online is low in absolute terms—with open cross-ideological exchanges and exposure to ideological diversity being fairly common (Bakshy, Messing, & Adamic, 2015). One possible explanation for the variability in results is that some studies have relied on self-selected samples of partisan individuals, whereas others have not (Bakshy et al., 2015; Conover et al., 2012; Westfall et al., 2015). Another is that some studies have excluded social-media platforms from their analysis (Gentzkow & Shapiro, 2011), despite the predominant role that social-networking sites play in online political communication. Our network-based estimates overcame

these problems and enabled us to investigate exposure to social-media messages about a wide set of issues. We expected to observe higher levels of ideological segregation in tweets (Twitter messages) related to political issues (or partisan identities) and lower levels in tweets that focused on other types of current events.

Some psychological perspectives suggest that liberals and conservatives would be equally likely to engage in selective exposure to information that confirms their preexisting opinions and would exhibit equivalent levels of ideological homophily, insofar as processes such as dissonance reduction, identity maintenance, and motivated reasoning are both highly general and prevalent in the general population (e.g., Hogg, 2007; Kahan, 2013; Munro et al., 2002). Other perspectives suggest that conservatives—because of heightened epistemic, existential, and relational needs to reduce uncertainty, threat, and social discord—would be more likely than liberals to prefer an echo-chamber environment (Jost, Glaser, Kruglanski, & Sulloway, 2003; Jost & Krochik, 2014; Stern, West, Jost, & Rule, 2014). Prior research focusing on traditional forms of media usage is inconclusive; some studies have obtained symmetrical patterns of selective exposure, dissonance avoidance, and ideological conformity (Iyengar & Hahn, 2009; Knobloch-Westerwick & Meng, 2009; Munro et al., 2002; Nisbet, Cooper, & Garrett, 2015), whereas others have suggested that conservatives are indeed more likely than liberals to exhibit these types of behaviors (Garrett, 2009b; Iyengar, Hahn, Krosnick, & Walker, 2008; Lau & Redlawsk, 2006; Mutz, 2006; Nam, Jost, & Van Bavel, 2013; Nyhan & Reifler, 2010; Sears & Freedman, 1967).

We analyzed “big data” sources to investigate both of these research questions, namely, (a) the extent to which the online media environment resembles an echo chamber characterized by selective exposure, ideological segregation, and political polarization or a “national conversation” in which individuals of differing ideological persuasions read and retweet one another’s messages and (b) whether liberals and conservatives behave similarly or dissimilarly in their use of social media when it comes to sharing information about politics and current events. By estimating the ideological preferences and social-network structures of 3.8 million Twitter users in the United States, we were able to examine patterns of public interaction and information diffusion with respect to 12 different political and nonpolitical issues.

Method

There is a limited but growing literature on how to infer social-media users’ attributes on the basis of their network positions. Although ideology is one of the key predictors of social and political behavior (Jost, 2006), only

a handful of studies have attempted to measure ideology using social-media data (e.g., Back et al., 2010; Conover et al., 2011; Kosinski et al., 2013). These studies have relied on self-selected samples and relatively crude operationalizations of ideology. Furthermore, previously used methods are not easily scalable to large social-media networks, which raises concerns about the validity of inferences drawn with respect to the entire population of users.

A latent space model of political ideology

Our approach builds on the theoretical logic of latent space models as applied to social-network data (Hoff et al., 2002). These models assume that the positions of individuals in an unobserved social space can be inferred on the basis of observed connections among them insofar as such connections are governed by the principle of homophily. In other words, if individuals are embedded in networks that are homogeneous with regard to sociodemographic and behavioral characteristics, then their connections can be interpreted as indications of similarity. There is reason to believe that social networks are indeed homophilous when it comes to political ideology (see Barberá, 2015; Bond & Messing, 2015).

Our statistical model assumes that the probability of a connection between user i and user j , both nodes of a given network, is negatively related to d_{ij} , their distance in a latent ideological space:

$$p(\mathbf{Y}_{ij} = 1 \mid \alpha_i, \beta_j, d_{ij}) = \text{Logistic}(\alpha_i + \beta_j - d_{ij}),$$

The model further assumes that the probability of a connection between i and j is independent of the probability of a connection between i and another user k , given user i 's and user j 's ideological positions and the random effects α_i and β_j , which account for the differences in their baseline probability of being connected to other users. By estimating d_{ij} , we obtain the relative positions of both user i and user j in the latent ideological space.

Latent space models are usually estimated using Markov-chain Monte Carlo methods, because standard maximum likelihood approaches are intractable for medium-to-large networks (Barberá, 2015; Shortreed, Handcock, & Hoff, 2006). However, this Bayesian approach becomes computationally inefficient for large-scale networks, such as those that are found on social-media sites. Therefore, we used correspondence analysis (Greenacre, 1993) to estimate latent parameters and reduce computational costs.¹ Correspondence analysis assumes that users are located in a multidimensional

space, and it computes users' positions by minimizing the distance with respect to political accounts using a weighted euclidean metric. The first step in this method is to standardize the matrix \mathbf{Y} by its column and row sums, which is equivalent to including fixed effects in the estimation (α_i and β_j ; see also Bonica, 2014). Then, one calculates the singular value decomposition of this new matrix, \mathbf{S} , which allows one to identify the plane closest to the data, where proximity is defined in terms of weighted least squares. Finally, users are projected onto this plane, with their coordinates being equivalent to their ideological positions.² These positions are then standardized to have a normal distribution with a mean of 0 and a standard deviation of 1, which facilitates the interpretation of the scale; a user with an estimated ideological position of -1 , for instance, will be located 1 standard deviation to the left of the average user.

To ensure that our estimation method located users in an ideological latent space, we initially restricted our matrix of connections to users' decisions about whom to "follow" regarding a subset of popular accounts with high ideological discrimination, such as accounts of the president, legislators, candidates for public office, media outlets, and interest groups. In Section 3.2 of the Supplemental Material available online, we summarize the results of a series of predictive checks, which demonstrated that the fit of our model was adequate; in other words, distance in the latent ideological space we identified was indeed predictive of the decisions users made about whom to follow.

Sampling of topics and users

In our analysis of online ideological polarization in the United States, we investigated the structure and content of Twitter conversations about 12 significant (political and nonpolitical) events and issues that arose from 2012 through 2014 (see Table 1). The political topics included issues that were discussed more frequently by Republicans than by Democrats (e.g., the U.S. federal budget) and issues that were discussed more frequently by Democrats than by Republicans (e.g., marriage equality, raising the minimum wage), as well as topics discussed equally by both (e.g., the 2012 presidential campaign, 2013 government shutdown, and 2014 State of the Union address). The nonpolitical topics ranged from sports and entertainment events (the 2014 Super Bowl, Winter Olympics, and Academy Awards) to mass tragedies (the Boston Marathon bombing, Newtown school shooting, and use of chemical weapons during the Syrian civil war). We identified a set of keywords for each topic and then used open-source tools developed by the Social Media and Political Participation (SMaPP) Lab (smapp.nyu.edu) at New York University (see SMaPP, n.d.) to collect a total of nearly

Table 1. Summary of the Tweet Collections on 12 Political and Nonpolitical Topics

Tweet collection	Period	Number of tweets (in millions)
2012 presidential campaign: obama, romney	8/15/2012–11/6/2012	62.3
2013 government shutdown: shutdown, #gopshutdown, boehner, furlough, . . .	10/1/2013–11/1/2013	12.4
Minimum wage: minimum wage, #raisethewage, #tenten, #timefor1010, . . .	2/3/2014–4/16/2014	0.2
Budget: budget, deficit, sequester, social security, . . .	6/1/2013–12/31/2013	7.7
Marriage equality: scotus, doma, gay marriage, same sex marriage, . . .	6/26/2013–12/2/2013	8.2
2014 State of the Union Address: state of the union, #sotu2014, republican response, . . .	1/27/2014–2/2/2014	2.7
Boston Marathon bombing: boston, marathon, explosion, #bostonmarathon	4/15/2013–4/30/2013	13.9
Newtown school shooting: newtown, sandy hook, gun control, #ctshooting, . . .	12/10/2012–1/8/2013	5.1
Syria: syria, intervention, chemical weapons, al-assad, . . .	8/28/2013–9/30/2013	7.8
2014 Super Bowl: super bowl, broncos, seahawks, touchdown, . . .	2/1/2014–2/3/2014	5.0
2014 Oscars: oscar, #oscar2014, academy awards, . . .	2/19/2014–3/11/2014	10.6
2014 Winter Olympic Games: olympics, sochi, team usa, #olympics2014, . . .	2/7/2014–2/19/2014	7.7

Note: The collections were based on searches for specific terms. Sample terms are listed in this table; the complete list of search terms used for all 12 collections is provided in Table S3 in the Supplemental Material.

150 million tweets that mentioned any of those keywords from the Twitter Streaming API (application program interface).

The sample of individuals in our study comprised all Twitter users who posted at least one tweet in this collection and followed at least one of the 1,206 political accounts we identified in the analysis described in the next section. We also applied simple activity, location, and spam filters (see Section 2.2 of the Supplemental Material) to try to ensure that all Twitter users included in our analysis corresponded to actual citizens (i.e., we excluded “ghost accounts” and “bots”). After applying these filters, we obtained a sample of 3.8 million active Twitter users in the United States.

Estimating and validating ideological placement

Our procedure for estimating ideological placement consisted of three stages (see Section 1 of the Supplemental Material for more details). First, we computed the ideological coordinates for users who followed at least 10 political accounts; this allowed us to identify the ideological latent space. Our initial list of political accounts included the Twitter accounts of the president and vice president, the Democratic and Republican parties, all members of Congress with more than 5,000 followers, and other relevant political figures (see Table S1 in the Supplemental Material for the complete list). Second, we expanded this list of political accounts to include other accounts that were commonly followed by liberal and conservative users in our sample (see Table S2 in the Supplemental Material). For instance, in this step, we added the accounts of Obama for America, *The Huffington Post*, Ready for Hillary, and Stephen Colbert (popular among liberals), as well as those of Rush Limbaugh,

Bobby Jindal, Tucker Carlson, and the Tea Party (popular among conservatives). Finally, we projected these additional accounts and their followers onto the same latent space to estimate the ideological positions of our entire sample of 3.8 million individuals, including the original political elites used to initiate the estimation process. The fact that the model estimated ideological positions for elites and nonelites was especially useful because it provided multiple options for validation.

As illustrated in Figure 1, we were indeed able to validate the ideology estimates using a range of offline measures. Figure 1a demonstrates that statewide medians of Twitter-based estimates of ideology in the United States were highly correlated with statewide averages of ideology calculated on the basis of survey and sociodemographic data (Lax & Phillips, 2012). Turning to elites, as shown in Figure 1b, we found that our ideological estimates for members of the House and Senate were highly correlated with a measure of ideology based on their roll-call votes in Congress, the current “gold standard” for estimating legislative ideology (Clinton, Jackman, & Rivers, 2004).³ Capitalizing on the public availability of party-registration data in some states, we matched a sample of Twitter accounts from California, Pennsylvania, Florida, Arkansas, and Ohio with voter-registration records (using full names and residential counties) and concluded that our Twitter-based ideal point estimates of ideology were good predictors of party registration, as indicated by the low degree of overlap in the distributions of ideological estimates for Democrats and Republicans, shown in Figure 1c. In particular, we found that setting a value of 0.5 in the latent ideological dimension as a threshold to distinguish registered Democrats from Republicans allowed us to correctly classify the party registration of 78% of matched voters. These tests confirmed that for

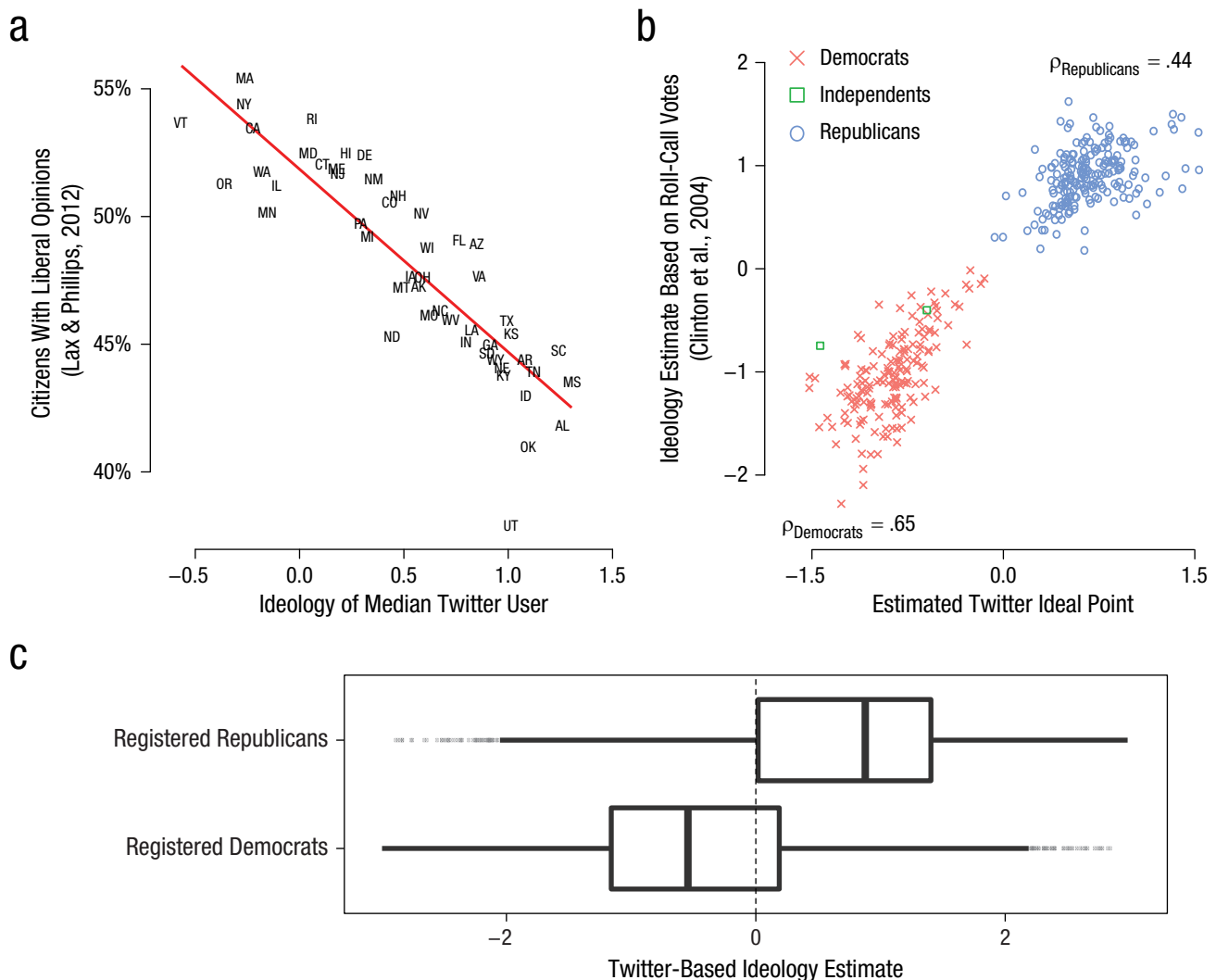


Fig. 1. Results of the tests validating the procedure for estimating ideological placement. The graph in (a) illustrates the correlation ($r = .87$) between the estimated ideological location of the median Twitter user in each state on a liberal-conservative latent scale and the percentage of citizens holding liberal opinions across different issues (estimated by Lax & Phillips, 2012, using a combination of survey and sociodemographic data). The graph in (b) illustrates the correlation ($r = .95$) between the ideal point estimates of the ideology of the 365 members of the 113th U.S. Congress with more than 5,000 followers, based on their Twitter follower networks, and the Congress members' ideology scores estimated from roll-call votes (Clinton et al., 2004). Within-party correlations are also shown. The graph in (c) displays the distribution of Twitter-based ideology estimates of a sample of Twitter users who were registered as Republicans and Democrats in California, Florida, Pennsylvania, Arkansas, and Ohio in 2012. For each group of voters, the box indicates the values of the 25th, 50th, and 75th percentiles of the ideology distribution, whereas the ends of the whiskers indicate the highest and lowest values within 1.5 times the interquartile range. Values beyond the end of the whiskers (outliers) are displayed as points. The dotted line indicates the position of the average voter (0, by construction).

the users in our sample, our method yielded ideology estimates that corresponded well to conventional measures of political ideology.

Results

Estimating ideological segregation and polarization

Our analysis focused on one of the most common forms of interaction on Twitter: retweeting, which involves

reposting another user's content with an attribution to the original author. Retweeting can be taken as an indicator of information diffusion, but not necessarily message endorsement (González-Bailón, Borge-Holthoefer, & Moreno, 2013). We utilized three different metrics of polarization: (a) the percentage of retweets that took place among individuals who were ideologically similar, (b) the degree of ideological homogeneity in communities of users detected in the retweet network, and (c) the average ideological extremity of the content that was spread via retweets.

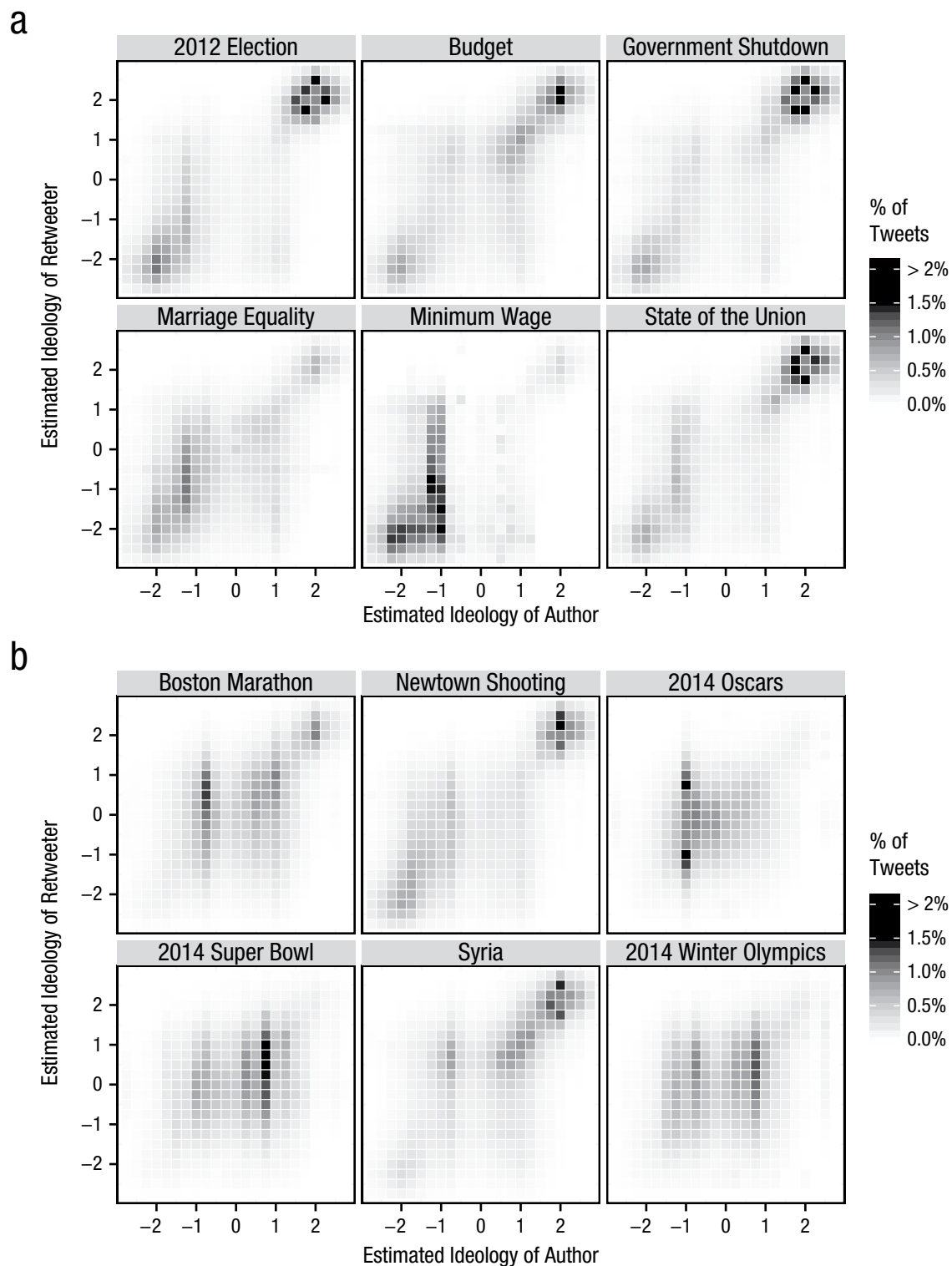


Fig. 2. Ideological polarization in retweeting of content on (a) the six political topics and (b) the six nonpolitical topics in the tweet collections. The intensity of the shading of each cell (of size $0.25\ SD \times 0.25\ SD$) represents the percentage of retweets that were published originally by users with estimated ideology of x and retweeted by users with estimated ideology of y . The highest level of polarization—which would occur if users spread only information transmitted by those who were ideologically indistinguishable from themselves—would correspond to 100% of retweets falling along the 45° line.

We found that polarization levels varied significantly as a function of time and topic. The heat maps in Figure 2

show the percentage of retweets on each of the 12 topics according to the estimated ideology of the tweets' original

authors and of the users who retweeted the messages. With respect to political issues such as the government shutdown and marriage equality, the vast majority of retweets occurred within ideological groups, as indicated by the shading clustered along the 45° lines in Figure 2a. That is, liberals tended to retweet tweets from other liberals, and conservatives were especially likely to retweet tweets from other conservatives. For example, 38% of all retweets concerning the 2012 election took place among extreme conservatives (more than 1 *SD* to the right of the mean), and 28% took place among extreme liberals (more than 1 *SD* to the left of the mean)—although each of these groups represented only 16% of the users in our sample.

At the same time, not all conversations were consistent with the traditional echo-chamber view of Internet communication (Garrett, 2009a). As Figure 2b shows, retweets about nonpolitical topics, such as the Boston Marathon bombing, the 2014 Super Bowl, and the 2014 Winter Olympics, surmounted ideological boundaries, as indicated by the lack of shading clustered around the 45° line in the heat maps, except at the far right for some topics. Thus, current events that are unrelated to politics do seem to have the capacity to stimulate discussion among individuals who differ in their political opinions. In other words, we found that ideological homophily in the propagation of content related to nonpolitical events is low; in this sense, discussions of current events do not strictly conform to the image of an echo chamber.

Figures 3a and 3b provide additional evidence of topic-specific dynamics in retweeting. For the collections pertaining to the 2012 presidential-election campaign and the 2014 Super Bowl, we treated each dyadic interaction as an edge (line) of a network in which each node (dot) was a social-media user. We then used a force-directed layout algorithm to generate graphic depictions of these large-scale networks and looked for clusters, or “cliques,” of users so that we could determine whether these groups were ideologically homogeneous. This algorithm placed users who retweeted each other’s tweets often very close to one another and placed users who rarely or never interacted further apart. The network of retweets concerning the election campaign (Fig. 3a) appears to have had two large and distinct ideological clusters, one composed of conservatives and the other of liberals. These groups of users were distant from one another because of the relatively low level of cross-ideological interaction overall. By contrast, the network of retweets concerning the Super Bowl does not display any large cluster that is clearly identifiable and ideologically homogeneous (Fig. 3b); instead, there are many small, heterogeneous clusters, which indicates that liberals and conservatives interacted with each other to a considerable degree.

Figure 3c summarizes the average ideological extremity of retweeted content, computed as the average absolute distance, for all retweets on each topic, between the original author and the ideological center. The obtained values indicate that discussions of topics such as the minimum wage and the government shutdown were highly polarized, with the most frequently dispersed content generated by a relatively extreme set of authors. In contrast, tweets about the Super Bowl and the Boston Marathon bombing, to take two examples, were more likely to be retweeted if they were generated by middle-of-the-road sources than if they were generated by more extreme sources.

Finally, we found that Twitter responses to some specific events—such as the tragic elementary-school shooting in Newtown, Connecticut—exhibited a dynamic shift from national conversation to echo chamber (see Fig. 3d, which shows the polarization index for five issues on a day-to-day basis). The initial pattern of information diffusion was intermediate, but the level of polarization increased over time, as the conversation shifted from the tragedy to a debate over gun-control policy. A similar pattern—in the opposite direction—was discernible in our collection of tweets related to the possibility of a military intervention in Syria. There was an initial increase in polarization as liberals and conservatives debated the issue, before attention to the topic subsided and the level of polarization dropped.

Estimating ideological asymmetry

To investigate whether there is an ideological asymmetry in segregation and polarization, we used Poisson regression models to estimate the baseline probability that one user would retweet another user’s post given the observed marginal rates of retweeting by liberals and conservatives as well as their likelihood of being retweeted. By computing these expected rates of retweeting, we could then examine whether cross-ideological retweets were more or less likely to occur than would be expected on the basis of chance (see Section 5 of the Supplemental Material for additional details). Figure 4 summarizes the results of this analysis.

Our analysis yielded three main results that are consistent with theory and research pertaining to ideological differences in epistemic, existential, and relational motivation. First, for all of the political topics, liberals were significantly more likely to engage in cross-ideological retweeting than were conservatives. For each retweet that took place between individuals of the same ideological orientation, there were between 0.20 and 0.30 cases of conservatives retweeting liberals, with some variation across topics, and between 0.25 and 0.50 cases of liberals retweeting conservatives. Second, the

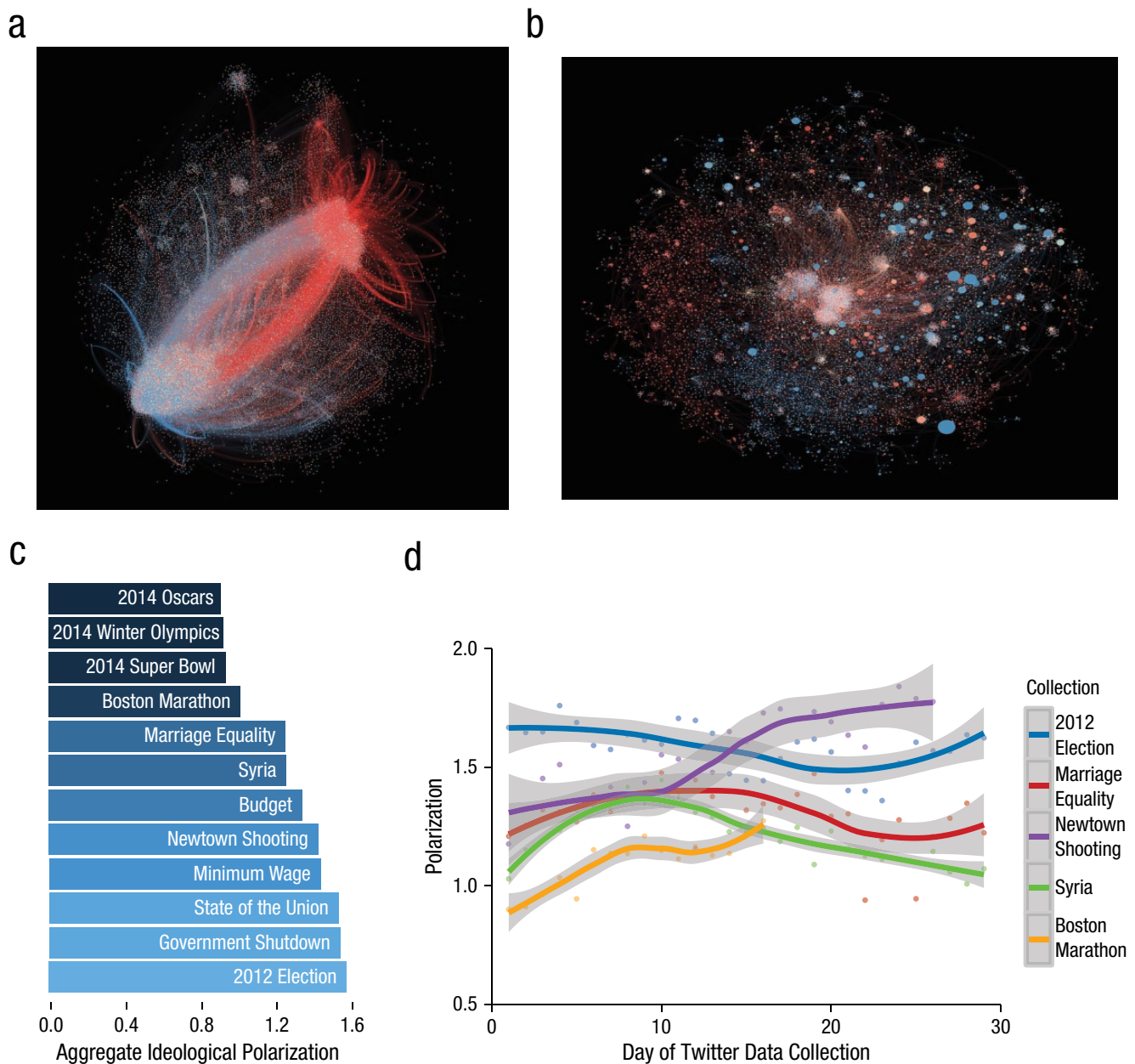


Fig. 3. Additional results on polarization in retweeting behavior. The graphics in (a) and (b), which were created using a force-directed layout algorithm, depict the retweet networks for the tweet collections on the 2012 election and the 2014 Super Bowl. Each node (dot) represents one user (from a random sample, weighted by activity), and each edge (line) represents a retweet. Nodes are colored according to the ideology estimate of the corresponding user, from very conservative (dark red) to very liberal (dark blue). Edges are colored according to the ideology estimate of the user whose tweet was retweeted. White color denotes areas with a large number of nodes whose placement in the figure overlap. The bar graph (c) displays the average level of political polarization for each of the 12 collections in our study. Polarization was calculated as the average absolute distance, for all retweets on a topic, between the original author and the ideological center. Higher levels of polarization imply that the information that was spread via retweets featured content that was more ideologically extreme. The graph in (d) illustrates the evolution of this index of polarization in information diffusion as a function of the number of days passed since each collection was started. Results are shown for a selection of five issues. Each data point indicates the estimated polarization index for a given day, and the curves correspond to local regression lines with loess smoothing, with 95% confidence intervals in gray.

rates of cross-ideological retweeting were generally higher for nonpolitical topics than for political topics, although in all cases they were lower than would be expected in the absence of ideological considerations

(i.e., to the left of the vertical dashed line in Fig. 4). Third, the ideological asymmetry in cross-ideological retweeting was generally smaller in magnitude for nonpolitical than for political topics. For the Boston

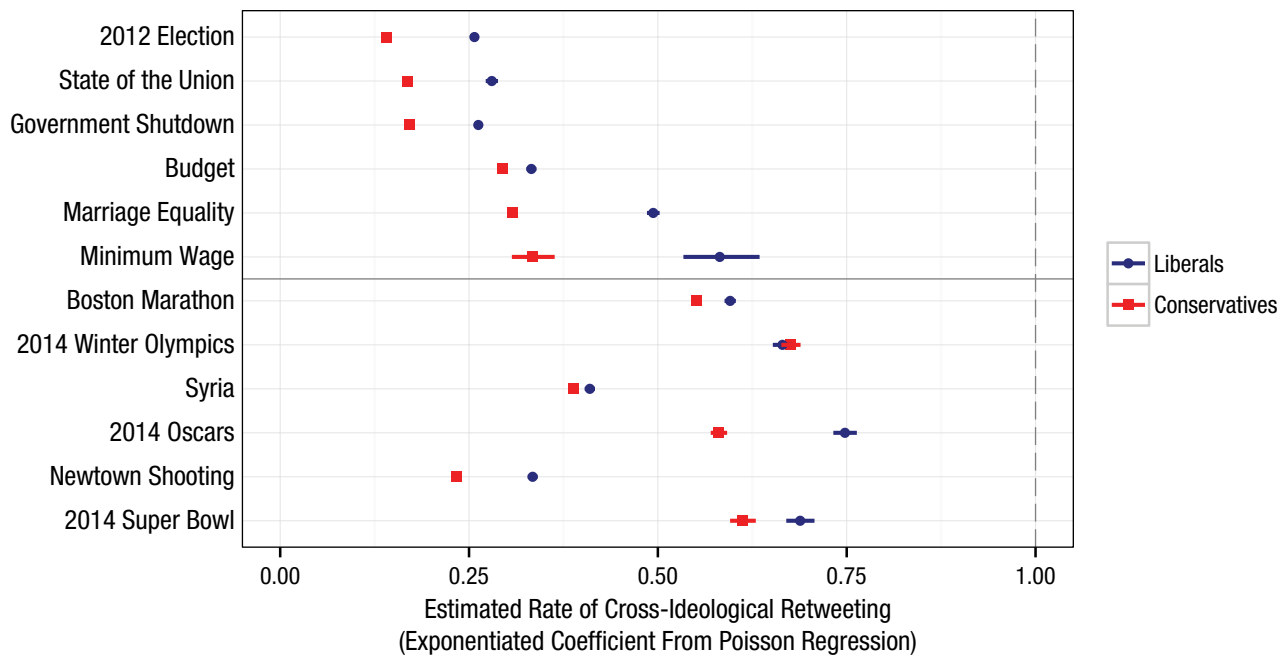


Fig. 4. Liberal-conservative asymmetries in cross-ideological retweeting. The graph shows the estimated rate of cross-ideological retweeting for each tweet collection and for each ideological group after adjusting for each group's propensity to retweet and be retweeted; each point corresponds to an exponentiated coefficient of a Poisson regression for the indicated topic and ideological group. The error bars indicate 99.9% confidence intervals (not visible in some cases because of their small size). An exponentiated coefficient of 1 (highlighted by the dashed vertical line) would indicate identical retweeting rates for individuals of the same and different ideological orientations—that is, a rate of cross-ideological retweeting that is equal to the rate of within-group retweeting.

Marathon bombing and the Super Bowl, the difference between liberals' and conservatives' rates of cross-ideological retweeting was small, and in the case of the Winter Olympics, the difference between liberals and conservatives was not statistically significant.

General Discussion

Overall, we observed that online communication structures are flexible and situation-specific, and that the aggregate level of political polarization depends heavily on the nature of the issue. For example, Twitter discussions of the 2012 election, the 2013 government shutdown, and the 2014 State of the Union Address resembled an echo chamber: Information about these events was exchanged primarily among individuals with highly similar ideological preferences. By contrast, responses to the Boston Marathon bombing in 2013, the 2014 Super Bowl, and the 2014 Winter Olympics fit the pattern of a national conversation, with individuals of differing ideological persuasions frequently reading and retweeting one another's messages. Public discussion of other issues, such as the aftermath of the Newtown shooting in 2012, reflected a dynamic process, beginning as national conversations but transforming fairly rapidly into highly polarized exchanges. With respect to both political and nonpolitical topics, liberals were more likely than conservatives to engage in cross-ideological retweeting. We do

not know the extent to which such behavior was carried out ironically or for the purpose of ideological criticism; in any case, it does seem that liberals were more likely than conservatives to expose themselves to differing opinions and to circulate those opinions.⁴

Our research has important implications for the study of social behavior, political communication, and democratic theory. First, we observed considerable variation across time periods and topics in the extent to which conversations on Twitter were politically polarized. This suggests that some previous studies may have overestimated the degree of mass political polarization. Our results reveal that homophilic tendencies in online interaction do not imply that information about current events is necessarily constrained by the walls of an echo chamber; in some cases, information diffusion permeates the entire network (see Bakshy et al., 2015, for a similar conclusion). Second, from a normative perspective, the fact that individuals receive news and information from diverse ideological sources may improve the quality of the informational environment, as well as the fidelity of political representation (Mutz, 2006). These findings highlight the rich potential of social-media platforms to capture the dynamics of public opinion, and they illustrate the value of using social-media data to examine questions about human behavior in naturally occurring social and political contexts.

Conclusion

The ability to estimate the ideological preferences of millions of Twitter users provides a historically unique opportunity to investigate political communication, ideological segregation, and political polarization, along with many other forms of social and political behavior. Our analysis revealed that communication structures are dynamic and flexible and that citizens do come into contact with information from diverse ideological perspectives, which is often assumed to redound to the benefit of social capital, political representation, and public policy. The popularization of social media as a means of communication within interpersonal networks is not inevitably bounded by ideological contours, especially when it comes to nonpolitical issues and events. When it comes to explicitly political issues, individuals are clearly more likely to pass on information that they have received from ideologically similar sources than to pass on information that they have received from dissimilar sources. At the same time, we observed that liberals are significantly more likely than conservatives to participate in cross-ideological dissemination of political and nonpolitical information. This suggests that an important and underappreciated ideological asymmetry may exist in the structure and function of online political communication.

Author Contributions

All of the authors contributed to the theorizing behind this study, to the research design, and to writing the manuscript. P. Barberá wrote the computer programs, conducted the research, analyzed the data, and prepared the tables, figures, and Supplemental Material.

Acknowledgments

The authors gratefully acknowledge the assistance of data scientists Duncan Penfold-Brown and Jonathan Ronen of New York University's Social Media and Political Participation (SMaPP) Lab, as well as comments on earlier versions of this article by Hal Arkes, Ken Benoit, Sam Gosling, Dustin Tingley, and several anonymous reviewers.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article. None of the funders were required to approve the publication of this article.

Funding

This research was supported by the INSPIRE program of the National Science Foundation (Award SES-1248077), the New York University Global Institute for Advanced Study, and Dean Thomas Carew's Research Investment Fund at New York University.

Supplemental Material

Additional supporting information can be found at <http://pss.sagepub.com/content/by/supplemental-data>

Open Practices



All data have been made publicly available through the Harvard Dataverse Network and can be accessed at <http://dx.doi.org/10.7910/DVN/F9ICHH>. All materials have been made publicly available at https://github.com/SMAPPNYU/echo_chambers. The complete Open Practices Disclosure for this article can be found at <http://pss.sagepub.com/content/by/supplemental-data>. This article has received badges for Open Data and Open Materials. More information about the Open Practices badges can be found at <https://osf.io/tvyxz/wiki/1.%20View%20the%20Badges/> and <http://pss.sagepub.com/content/25/1/3.full>.

Notes

1. Figure S2 in the Supplemental Material available online demonstrates that this approach yields estimates that are highly correlated with those computed using Markov-chain Monte Carlo methods.
2. See Section 1 of the Supplemental Material for additional technical details on the estimation method.
3. Legislators' positions were placed on a latent ideological scale on the basis of their observed voting behavior. These locations were estimated by assuming that in their voting decisions for each bill, legislators minimized the distance between their positions and that of the "yea" or "nay" option they chose.
4. Several social and psychological factors—including age and personality differences between liberals and conservatives—could help to explain the ideological asymmetry we observed. Unfortunately, data on these variables were not available for all the Twitter users in our sample. We explored possible effects of age by relying on a subsample of Twitter users that was matched with voter-registration files (see Section 5.3 of the Supplemental Material). We observed that for all age groups, liberals were more likely than conservatives to engage in cross-ideological retweeting in the political and nonpolitical domains, with one exception: Young conservatives (ages 18–24) were just as likely as young liberals to retweet political messages across ideological boundaries. Examining personality characteristics, we computed aggregate-level correlations (on a state-wide basis) and observed that states with higher average scores on openness tend to be more liberal and to have more cross-ideological retweeting behavior, whereas states with higher average scores on conscientiousness tend to be more conservative and to have less cross-ideological retweeting behavior.

References

- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36–43). doi:10.1145/1134271.1134277
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect

- actual personality, not self-idealization. *Psychological Science*, 21, 372–374. doi:10.1177/0956797609360756
- Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348, 1130–1132. doi:10.1126/science.aaa1160
- Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, 23, 76–91. doi:10.1093/pan/mpu011
- Berra, Y. (2008). *You can observe a lot by watching: What I've learned about teamwork from the Yankees and life*. Hoboken, NJ: Wiley.
- Bond, R., & Messing, S. (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review*, 109, 62–78.
- Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, 58, 367–386. doi:10.1111/ajps.12062
- Clinton, J. S., Jackman, S., & Rivers, D. (2004). The statistical analysis of roll call data. *American Political Science Review*, 98, 355–370.
- Conover, M. D., Gonçalves, B., Flammini, A., & Menczer, F. (2012). Partisan asymmetries in online political activity. *EPJ Data Science*, 1, 1–19. doi:10.1140/epjds6
- Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Political polarization on Twitter. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (pp. 89–96). Retrieved from <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2847/3275>
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 86, 345–355.
- Garrett, R. K. (2009a). Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14, 265–285. doi:10.1111/j.1083-6101.2009.01440.x
- Garrett, R. K. (2009b). Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, 59, 676–699. doi:10.1111/j.1460-2466.2009.01452.x
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, 126, 1799–1839. doi:10.3386/w15916
- Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, 333, 1878–1881. doi:10.1126/science.1202775
- González-Bailón, S., Borge-Holthoefer, J., & Moreno, Y. (2013). Broadcasters and hidden influentials in online protest diffusion. *American Behavioral Scientist*, 57, 943–965. doi:10.1177/0002764213479371
- Greenacre, M. J. (1993). *Correspondence analysis in practice*. London, England: Academic Press.
- Hoff, P., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97, 1090–1098. doi:10.1198/016214502388618906
- Hogg, M. A. (2007). Uncertainty-identity theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 39, pp. 69–126). San Diego, CA: Academic Press. doi:10.1016/S0065-2601(06)39002-8
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, USA*, 101, 16385–16389. doi:10.1073/pnas.0403723101
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59, 19–39. doi:10.1111/j.1460-2466.2008.01402.x
- Iyengar, S., Hahn, K. S., Krosnick, J. A., & Walker, J. (2008). Selective exposure to campaign communication: The role of anticipated agreement and issue public membership. *Journal of Politics*, 70, 186–200. doi:10.1017/s0022381607080139
- Jost, J. T. (2006). The end of the end of ideology. *American Psychologist*, 61, 651–670. doi:10.1037/0003-066X.61.7.651
- Jost, J. T., Glaser, J., Kruglanski, A. W., & Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375. doi:10.1037/0033-2909.129.3.339
- Jost, J. T., & Krochik, M. (2014). Ideological differences in epistemic motivation: Implications for attitude structure, depth of information processing, susceptibility to persuasion, and stereotyping. *Advances in Motivation Science*, 1, 181–231. doi:10.1016/bs.adms.2014.08.005
- Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8, 407–424. doi:10.2139/ssrn.2182588
- Knobloch-Westerwick, S., & Meng, J. (2009). Looking the other way: Selective exposure to attitude-consistent and counter-attitudinal political information. *Communication Research*, 36, 426–448. doi:10.1177/0093650209333030
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, USA*, 110, 5802–5805. doi:10.1073/pnas.1218772110
- Lau, R. R., & Redlawsk, D. P. (2006). *How voters decide: Information processing during election campaigns*. Cambridge, England: Cambridge University Press.
- Lax, J. R., & Phillips, J. H. (2012). The democratic deficit in the states. *American Journal of Political Science*, 56, 148–166. doi:10.1111/j.1540-5907.2011.00537.x
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 28, 415–444. doi:10.1146/annurev.soc.27.1.415
- Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased assimilation of socio-political arguments: Evaluating the 1996 U.S. presidential debate. *Basic and Applied Social Psychology*, 24, 15–26. doi:10.1207/s15324834basps2401_2
- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. New York, NY: Cambridge University Press.

- Nam, H. H., Jost, J. T., & Van Bavel, J. J. (2013). "Not for all the tea in China!" Political ideology and the avoidance of dissonance-arousing situations. *PLoS ONE*, 8(4), Article e59837. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0059837>
- Nisbet, E. C., Cooper, K. E., & Garrett, R. K. (2015). The partisan brain: How dissonant science messages lead conservatives and liberals to (dis)trust science. *The ANNALS of the American Academy of Political and Social Science*, 658, 36–66. doi:10.1177/0002716214555474
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32, 303–330. doi:10.1007/s11109-010-9112-2
- Pfau, M., Houston, J. B., & Semmler, S. M. (2007). *Mediating the vote: The changing media landscape in U.S. presidential campaigns*. New York, NY: Rowman & Littlefield.
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. New York, NY: Cambridge University Press.
- Sears, D. O., & Freedman, J. L. (1967). Selective exposure to information: A critical review. *Public Opinion Quarterly*, 31, 194–213.
- Shortreed, S., Handcock, M. S., & Hoff, P. (2006). Positional estimation within a latent space model for networks. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2, 24–33. doi:10.1027/1614-2241.2.1.24
- Social Media and Political Participation. (n.d.). *Data collection and analysis tools*. Retrieved from <http://smapp.nyu.edu/toolsNcode.html>
- Stern, C., West, T. V., Jost, J. T., & Rule, N. O. (2014). "Ditto heads": Do conservatives perceive greater consensus within their ranks than liberals? *Personality and Social Psychology Bulletin*, 40, 1162–1177.
- Van Boven, L., Judd, C. M., & Sherman, D. K. (2012). Political polarization projection: Social projection of partisan attitude extremity and attitudinal processes. *Journal of Personality and Social Psychology*, 103, 84–100.
- Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Non-reactive research in the social sciences*. Skokie, IL: Rand McNally.
- Westfall, J., Van Boven, L., Chambers, J., & Judd, C. M. (2015). Perceiving political polarization in the United States: Party identity strength and attitude extremity exacerbate the perceived partisan divide. *Perspectives on Psychological Science*, 10, 145–158.