

# Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook

ROBERT BOND *Ohio State University*

SOLOMON MESSING *Facebook Data Science*

**W**e demonstrate that social media data represent a useful resource for testing models of legislative and individual-level political behavior and attitudes. First, we develop a model to estimate the ideology of politicians and their supporters using social media data on individual citizens' endorsements of political figures. Our measure allows us to place politicians and more than 6 million citizens who are active in social media on the same metric. We validate the ideological estimates that result from the scaling process by showing they correlate highly with existing measures of ideology from Congress, and with individual-level self-reported political views. Finally, we use these measures to study the relationship between ideology and age, social relationships and ideology, and the relationship between friend ideology and turnout.

**M**any theories in political science rely on ideology at their core, whether they are explanations for individual behavior and preferences, governmental relations, or links between them. However, ideology has proven difficult to explicate and measure, in large part because it is impossible to directly observe: we can only examine indicators such as responses to survey questions, political donations, votes, and judicial decisions. One problem with this patchwork of indicator measures is the difficulty of studying ideology across domains. Although we have established reliable techniques for measuring ideology among individuals and legislators, such as survey measures and roll-call vote analysis, methods for jointly estimating the ideologies of ordinary citizens and elite actors have only recently been developed.

To understand the relationship between elite ideology and beliefs of ordinary citizens, we need measures of ideology that allow us to place ordinary citizens and elites in the same ideological space. For example, a longstanding debate in political science concerns whether the American public has become more ideologically polarized in the last 40 years (e.g., Abramowitz and Saunders 2008; Fiorina, Abrams, and Pope 2006). If so, does mass polarization drive elite polarization, or vice versa? To test these theories, we need joint ideology measures that put elite actors and ordinary citizens on the same scale.

The lack of comparable ideological estimates for individuals and elites can be attributed in part to a paucity

of suitable data. Although scholars continue to develop methods that use text as data to measure ideology (Laver, Benoit, and Garry 2003; Monroe, Colaresi, and Quinn 2009; Monroe and Maeda 2004), text data have problems that are yet unsolved—in addition to problems related to human language's high dimensionality, more fundamental problems can confound such efforts: not only are data on how ordinary citizens talk about issues related to ideology sparse, but also the context of communication for elites and ordinary citizens often differs, and each may use different language to describe the same underlying ideological phenomena (polysemy), or use the same language to describe different things (synonymy). While complex data such as human language may one day provide a superior measure of something as complex as ideology, further tools need to be developed to address these issues.

The most promising avenues for estimating ideology comprise discrete behavioral data that reveal preferences, including roll-call votes (Poole and Rosenthal 1997), cosponsorship records (Aleman et al. 2009), campaign finance contributions (Bonica 2013), and—as in this article—expressions of support for elites by ordinary citizens. Previously, obtaining sufficiently large behavioral datasets about ordinary citizens' political preferences has been cost prohibitive. Using large data sources, such as campaign contributions and online data, we are now able to use methods similar to those used on political elites to estimate the ideology of a broader set of political actors. These large-scale data sources allow researchers to view ideology at a more “macro” level, with the ability to couple the ideological estimates produced with other data sources. Using vast, emerging data sets like these allows political scientists to study ideology from new perspectives (Lazer et al. 2009).

Although methods for jointly scaling elites and masses using campaign finance records are promising (Bonica 2013), they should constitute the beginning of our efforts to produce measures that allow us to place elites and ordinary citizens on the same scale. While estimates based on campaign finance (or CFscores) extend available data beyond law-makers, they are

---

Robert Bond is Assistant Professor, School of Communication, Ohio State University, 3072 Derby Hall 154 North Oval Mall, Columbus, OH 43210 ([bond.136@osu.edu](mailto:bond.136@osu.edu)).

Solomon Messing is Research Scientist, Facebook Data Science, 1601 Willow Rd, Menlo Park, CA 94025, ([solomon@fb.com](mailto:solomon@fb.com)).

This paper was presented at the 2012 meeting of the Midwest Political Science Association, the 2012 Cal APSA meeting, and the 2012 Human Nature Group Retreat. We would like to thank the participants at these seminars as well as Scott Desposato, Chris Fariss, Will Fithian, James Fowler, Justin Grimmer, Alex Hughes, Gary Jacobson, Jason Jones, Thad Kousser, Michael Rivera, Jaime Settle, Neil Visalvanich, four anonymous reviewers, and the editors at the *American Political Science Review* for helpful comments and suggestions.

restricted to a particular type of ideological expression—giving money to campaigns, a behavior limited to class of individuals who have substantial financial resources and who choose to express their ideological preferences with money. Only donors who give at least \$250 to a particular campaign are named in campaign finance reports, though many give to campaigns in much smaller amounts—46% of Barack Obama’s 2008 presidential campaign donations were \$200 or less (Malbin 2009). While CFscores extends our ability to estimate ideological positions from politicians to PACs, the citizen-level estimates it produces cover an elite class of private donors who give large amounts to campaigns. Our ability to use such estimates to draw inferences about the ideological preferences for the voting public at large is limited.

In this article, we contribute to the effort to jointly measure elite and mass ideology using social media data. In contrast to donations, candidate endorsements in social media do not entail any financial outlay; any individual with an account may endorse any political entity she wishes. Furthermore, individuals can endorse any political entity that has a presence on the platform, meaning that such data allow for joint estimation of ideology for individuals and a broad range of political actors. This ideology measure is not fully representative—individuals who endorse political entities in social media have higher levels of political interest than average citizens, while individuals who are active in social media generally have higher levels of education and income. Nonetheless, this approach offers significant advantages. For instance, individuals in this data set need not overcome costly structural barriers to express their support of political candidates, nor do they comprise a limited, homogenous group of donors—instead they are drawn from a much broader swath of society. The potential generalizability of these data provides a compelling reason to examine and validate the ideological estimates it produces.

We proceed as follows: first we review the extant literature on methods for measuring ideology, both in populations of elites and for ordinary citizens. Following that, we explain how our social media endorsement data are structured and the how they are applicable to estimating ideology. Next, we describe our estimation technique and the resulting estimates. We then validate our estimates against other measures of ideology for both the elite and mass populations. The next sections apply the measure, investigating the relationship between age and ideology, ideology’s structure in social networks, and whether having friends with more distant ideologies decreases the likelihood of voting. Last, we conclude with a discussion of how this technique opens directions in the study of ideology and suggest future work using these data.

## MEASUREMENT OF IDEOLOGY

Political scientists have generally measured ideology by asking individuals to place themselves on a 7-point

liberal-conservative scale.<sup>1</sup> Although this measure’s ubiquity allows us to compare ideology across studies and across time, recent work has shown that it has a number of methodological problems. It cannot account for the multidimensionality of ideology (Feldman 1998; Treier and Hillygus 2009)—respondents might place themselves on a social, political, or economic ideology scale, or perhaps some combination thereof. The measure is recorded as an interval rather than a continuous measure (Jackson 1983; Kinder 1983), making it difficult to analyze with conventional statistical techniques. The measure also lacks reliability. Survey respondents may interpret the question differently, particularly respondents with liberal ideologies who are hesitant to be labeled “liberal” (Ansolabehere, Rodden, and Snyder 2008; Schiffer 2000). Although these issues are not so problematic that the measure should be abandoned, and researchers have proposed methods for addressing these concerns (Wood and Oliver 2012), they suggest that other measures should be explored.

Techniques to measure the ideology of political elites frequently use the choices that elites make as data from which an ideology measure may be deduced. The most well known of these techniques is the NOMINATE method of ideal point estimation using roll-call votes for members of Congress (Poole and Rosenthal 1997). Methods of ideal point estimation have been examined further in Congress (Clinton, Jackman, and Rivers 2004), and extended to other types of elite actors, such as members of the Supreme Court (Martin and Quinn 2002) and European Parliament (Hix, Noury, and Roland 2006). These techniques have provided researchers with estimates of the ideal points of actors within a given institution, but have largely left researchers without tools to compare the ideology of actors across them.

Typically, one cannot estimate the ideologies of a diverse set of political actors because they make disjoint choices. In order to jointly estimate the ideologies of actors with (primarily) disjoint choice sets one must have some set of choices that “bridges” choice divides (Bafumi and Herron 2010; Bailey 2007; Gerber and Lewis 2004; Poole and Rosenthal 1997). As with campaign contributions, Facebook’s data on which users support which candidates represent the bridge actors necessary to jointly estimate the ideology of politicians and ordinary citizens. Users are able to “like” pages regardless of their political institution, meaning that all pages are part of the same choice set. Bridge actors, such as Facebook users or donors, serve two important purposes. First, they bridge actors that otherwise would not be connected, such as members of Congress and

<sup>1</sup> Although question wording varies, a common version is, “We hear a lot of talk these days about liberals and conservatives. Here is a 7-point scale on which the political views that people might hold are arranged from extremely liberal to extremely conservative. Where would you place yourself on this scale, or haven’t you thought much about this?” Respondents are then given the choice to identify themselves as liberal or conservative, “extremely” or “slightly” liberal or conservative, “moderate/middle of the road,” or “don’t know/haven’t thought about it much.”

mayors of cities. Second, they span the divide between politicians and individuals.

To put elite political actors and ordinary citizens on the same ideological scale we use data from Facebook that mitigate limitations that previous measures of ideology faced in terms of bridging diverse sets of elite actors and including measures of the ideology of ordinary citizens. For Facebook users, showing support for political figures is simple, relatively costless, and requires little cognitive effort. Our approach uses singular value decomposition (SVD) on the transformed matrix of user to political page connections on Facebook to estimate the ideological positions of Facebook users and the political pages they support. These estimates are consistent with the first ideological dimension recovered from roll-call data (Clinton, Jackman, and Rivers 2004; Poole and Rosenthal 1997) and with individuals' self-reported political views indicated on the user's Facebook profile page.

## SOCIAL MEDIA ENDORSEMENT DATA

Social media serves not only as a platform to communicate with social contacts and share digital media, but also as a forum for political communication. As more individuals utilize this forum, they create rich data sources that can be used to understand political phenomena.<sup>2</sup> Political scientists have begun to study how people engage and express their political viewpoints online using this rich data source (e.g., Barberá 2014, Bond et al. 2012; Butler and Broockman 2011; Butler, Karpowitz, and Pope 2012; Grimmer, Messing, and Westwood 2012; Messing and Westwood 2012; Mutz and Young 2011; Wojcieszak and Mutz 2009).

Social media enable ordinary citizens to endorse and communicate with political figures and elites. On Facebook, in addition to displaying demographic, educational, and professional information in their profile, users can "like" pages associated with figures such as politicians, celebrities, musicians, television shows, books, movies, etc. Users do so by listing such entities on their profile, by visiting a politician's page on Facebook, or they may be prompted to like pages by friends. Liking political figures may be communicated to the user's social community via the political figure's page, the user's page, and the "news feed" homepage that friends of the user see. Liking pages makes the user a "fan" of that page, meaning that the individual may see content published on the page in their news feed. Facebook also maintains an accounting of pages that are "official."<sup>3</sup> We scale ideology for both users

<sup>2</sup> More than half of Americans use social media on a regular basis (Facebook 2011) and American Internet users spent more time on Facebook than any other single Internet destination in 2011 by nearly two orders of magnitude according to Nielsen, with the average person spending 7–8 hours per month (Nielsen 2011). Over 42 percent of Americans reported learning something about the 2012 campaign on Facebook, according to Pew (2012).

<sup>3</sup> For instance, there are many pages that are about President Obama in one way or another, but only one page ([www.facebook.com/barackobama](http://www.facebook.com/barackobama)) is official and is ostensibly maintained by the President.

and political figures using endorsements of official pages.

Facebook is an especially appropriate platform to examine the suitability of social media data to generate joint ideological estimates. On Facebook, users self-report demographic and political data, making it possible to cross-reference these estimates with self-reported ideological measures, and examine variation by demographic category. We can validate legislators' ideological estimates based on traditional measures such as DW-NOMINATE. Hence, our method can be extended to other social networking sites where public data on the politicians individuals endorse and follow are readily available to researchers, including Twitter. However, because Facebook collects demographic data, including self-reported political affiliation, it serves as a more appropriate platform to validate this approach.

For most results below, we used data collected on March 1, 2011 from all U.S. users of Facebook over age 18 who had publicly liked at least two of the official political pages on Facebook. This constituted 6.2 million individuals and 1,223 pages. Only official pages identified by Facebook were included. We also collected data from March 1, 2012 to see for whom ideology changed year over year. Finally, we collected data from November 2, 2010, the day of the Congressional election that year, in order to calculate an ideology score that we use to study its relationship to voting, as explained below. All data were de-identified.

## USING SOCIAL MEDIA DATA TO SCALE IDEOLOGICAL POSITIONS

Unlike roll-call data that are ideal for scaling, social media endorsements, much like campaign contribution data, require some processing prior to scaling. Roll-call data are well suited for scaling because votes are coded as either "yea" or "nay," and abstentions may be simply treated as missing data. For both Facebook and campaign contribution data, the presence of relationships is clear, but the absence of a relationship is ambiguous.<sup>4</sup> The lack of a supporting relationship may be related to ideological considerations, lack of knowledge about the candidate, or an unwillingness to make their supportive relationships public. As with campaign contribution data, Facebook users may choose to support any combination of political figures. Although this is also the case for campaign contributions, giving to candidates is influenced by an individual's budget, which may preclude an individual from giving to the full set of candidates she supports, or from giving at all. Nothing precludes a Facebook user from supporting a candidate and her opponent. One approach to simplify the analysis treats the data as choices between incumbent-challenger pairs. However, many of the political figures

<sup>4</sup> The fact that the data include only the presence of a relationship is not unlike other data sources that have been used to scale the ideology of political actors, such as legislative cosponsorships (Aleman et al. 2009; Talbert and Potoski 2009) and campaign finance records (Bonica 2013).

we wish to scale run for office against minimal opposition or do not run for office at all, meaning that their opposition has few or no supporters.

### Model of Endorsement

Suppose  $n$  users choose whether to like  $m$  candidates. Each user  $i = 1, \dots, n$  chooses whether to like candidate  $j = 1, \dots, m$  by comparing the candidate's position at  $\zeta_j$  and the status quo (in this case, not publicly supporting the candidate) located at  $\psi_j$ , both in  $\mathbf{R}^d$ , where  $d =$  dimensions of ideology space. Let

$$y_{ij} = \begin{cases} 1 & \text{if user } i \text{ likes candidate } j, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

User  $i$  receives utility for supporting candidates close to her own ideal point  $\mathbf{x}_i$  in  $\mathbf{R}^d$  policy space. We can specify this ideological utility as a quadratic loss function that depends on the location of the candidate and the status quo:

$$\begin{aligned} U_{ij}^{\text{candidate}} &= -\|\mathbf{x}_i - \zeta_j\|^2, \\ U_{ij}^{\text{status quo}} &= -\|\mathbf{x}_i - \psi_j\|^2. \end{aligned} \quad (2)$$

The net benefit of liking is then the difference in these two utilities,

$$\begin{aligned} U_{ij}^{\text{like}} &= U_{ij}^{\text{candidate}} - U_{ij}^{\text{status quo}} \\ &= -\|\mathbf{x}_i - \zeta_j\|^2 + \|\mathbf{x}_i - \psi_j\|^2. \end{aligned} \quad (3)$$

Notice that the utility of liking is *decreasing* in distance between the candidate and the user, but *increasing* in the distance between the status quo and the user.

Finally, suppose also that the utility of liking is increased by candidate-specific factors  $\phi_j$  that govern how desirable each political page is (some pages are more popular and, perhaps, easier to find on the site) and user-specific factors  $\eta_i$  that govern each user's propensity to support candidates (some users get greater utility from the act of liking, and are thus more likely to engage in the activity than others). Putting these together with liking utility yields

$$U_{ij} = -\|\mathbf{x}_i - \zeta_j\|^2 + \|\mathbf{x}_i - \psi_j\|^2 + \eta_i + \phi_j. \quad (4)$$

To group row and column terms into new variables, we let  $\beta_j = \psi_j - \zeta_j$ , and  $\theta_j = \psi_j^2 - \zeta_j^2 + \phi_j$ . Thus (4) simplifies to

$$U_{ij} = -2\mathbf{x}_i\beta_j + \eta_i + \theta_j. \quad (5)$$

We do not observe direct utilities, but we do observe likes. Suppose that observing a like means that the true utility of endorsing is high, while not observing one means that the true utility is low (without loss of generality, suppose the utilities are 1 and 0, respectively).

Not all likes yield exactly the same utility, so we can think of the true utility as being equal to a function of the observed like ( $y_{ij}$ ) minus an error term ( $v_{ij}$ ):

$$U_{ij} = y_{ij} - v_{ij} \quad (6)$$

Substituting, we get

$$y_{ij} = -2\mathbf{x}_i\beta_j + \eta_i + \theta_j + v_{ij}. \quad (7)$$

To further simplify the model, we factor out the  $\eta$  and  $\theta$  terms by employing the *double-center operator*  $D(\cdot)$  defined for a matrix  $\mathbf{Z}$  to be each element minus its row and column means plus its grand mean divided by  $-2$ :

$$D(z_{ij}) = (z_{ij} - \bar{z}_i - \bar{z}_j + \bar{z}_{..})/(-2). \quad (8)$$

In the literature that utilizes roll-call votes to estimate ideology, Poole (2005) and Clinton, Jackman, and Rivers (2004) discuss the use of the double-center operator on a *squared distance* matrix, not on the roll-call matrix itself. The effect of this operator is to generate a new matrix with all row and column means equal to zero. As a result, any term that does not interact with *both* a row and column variable will factor out of the matrix.

Suppose  $v_{ij}$  is an independent and identically distributed random variable drawn from a stable density. Suppose further, without loss of generality, that the dimension-by-dimension means of  $\mathbf{x}$  and  $\beta$  equal 0. If so, then applying the double-center operator in equation (8) to both sides of equation (7) yields

$$D(y_{ij}) = \mathbf{x}_i\beta_j + \epsilon_{ij}, \quad (9)$$

where the new error term  $\epsilon_{ij}$  is also a stable density defining the stochastic component of the identity. We can now use singular value decomposition (SVD) of the double-center matrix of likes to find the best  $d$  dimensional approximation of  $\mathbf{x}_i$  and  $\beta_j$  (Eckart and Young 1936):

$$D(\mathbf{Y}) = \mathbf{X}\mathbf{\Sigma}\mathbf{B}, \quad (10)$$

where  $\mathbf{Y}$  is the observed matrix of likes,  $\mathbf{X}$  is an  $n \times n$  matrix of user ideology locations,  $\mathbf{\Sigma}$  is a  $n \times m$  matrix with a diagonal of singular values, and  $\mathbf{B}$  is a  $m \times m$  matrix of  $\beta$ s. The  $d$  largest singular values correspond to the  $d$  columns of  $\mathbf{X}$  and  $d$  rows of  $\mathbf{B}$  that generate the best fitting estimates of  $\mathbf{x}_i$  and  $\beta_j$  (Eckart and Young 1936).

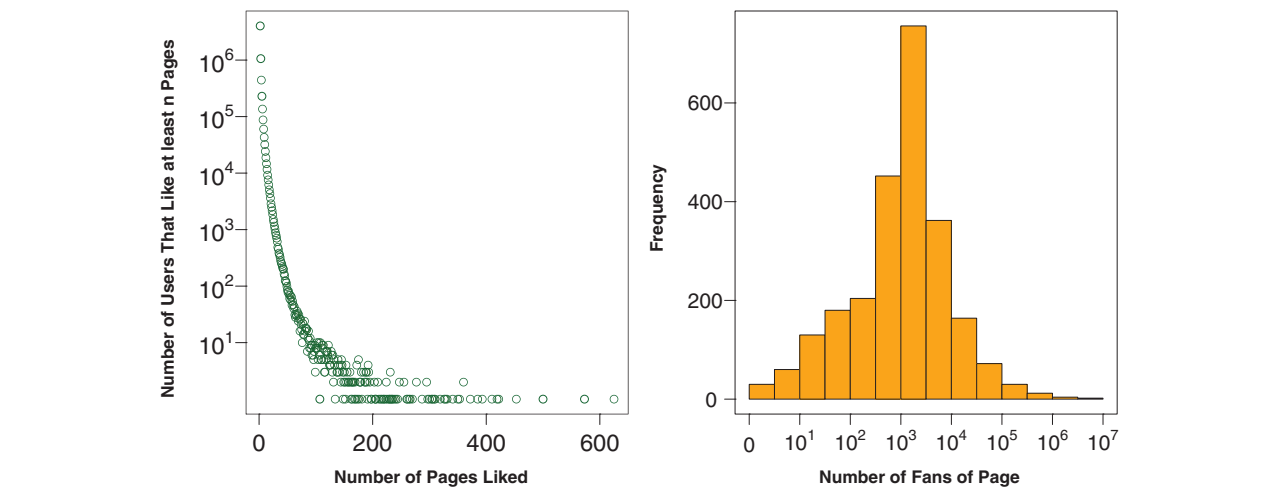
Although it is possible to analyze the full matrix of likes from users to political pages, the candidate ( $\phi_j$ ) and user ( $\eta_i$ ) specific factors mentioned above bias the estimation. Because some pages are so much more popular than others, and to a lesser extent because some users like many more politicians than average, the estimation yields ideological estimates that are weighted

**TABLE 1. The First Ten Rows of the User by Political Page Matrix**

	user 1	user 2	user 3	user 4	user 5	user 6	user 6	user 8	user 9	user 10
Barack Obama	1	1	1	.	.	.	.	.	.	.
Mitt Romney	1	.	.	.	.	.	.	1	.	.
Howard Dean	1	.	.	.	.	.	.	.	.	1
Joe Biden	.	.	.	.	.	.	.	.	.	.
Mike Bloomberg	1	.	1	.	.	.	.	.	.	.
Anthony Weiner	1	.	.	.	.	.	.	.	.	.
Deval Patrick	1	.	.	.	.	.	.	.	.	.
Diane Feinstein	1	.	1	.	.	.	.	.	1	.
Sarah Palin	.	.	.	.	.	.	.	.	.	.
Nancy Pelosi	.	.	.	.	.	.	1	.	.	.

Note: Entries in the matrix are dichotomous, where 1 means that the user has liked the page and . means that the user has not.

**FIGURE 1. The Left Panel Shows the Distribution of the Number of Pages that Each User Likes; the Right Panel Shows the Distribution of the Number of Fans that Each Page Has**



by the relative popularity of the candidates. For instance, Obama is to the *extreme* left of the distribution and Romney and Palin are to the *extreme* right, while candidates who have few likes are in the middle of the distribution regardless of their ideological views.

To account for political page- and user-specific factors, we compute a political page adjacency matrix,  $A = XX'$ , in which the rows and columns correspond to political pages and entries consist of the number of times an individual user likes both pages, and then compute the ratio of common users, corresponding to an agreement matrix,  $G = a_{ij} / \text{diag}(a_i)$ , described in further detail below. We then apply SVD to the  $G$  matrix to estimate  $x_i$ .

### Estimation of Ideology from Endorsements

We begin by creating a matrix in which each column represents a user and each row represents an official Facebook page about politics. We limit our data to Facebook users in the United States who are over age

18 who endorse, or “like,” at least two political pages.<sup>5</sup> This leaves us with approximately 6.2 million users and 1,223 pages and 18 million like actions from users to pages. An example of the first ten rows and first ten columns of the bipartite matrix is in Table 1.

A few things should be clear from this example and the summary statistics described here. First, there are few likes relative to the size of the matrix overall, making the matrix sparse. Second, users vary in the number of candidates they support. Although we limit the data to include users who like at a minimum two candidates’ pages, the average number of likes is 3.04 and the maximum number of pages liked is 625. Figure 1 shows the full distribution of the number of likes for users and the number of fans per page. Most users like only a few pages, but a few like many. Similarly, political pages vary in the amount of support from users they attract.

<sup>5</sup> We exclude users who like only one page and pages with only one supporter as they do not add any additional information to the matrix. These users would only be counted in the diagonals of the affiliation matrix described below.

**TABLE 2. The First Ten Rows of the Affiliation Matrix**

	Obama	Romney	Dean	Biden	Bloomberg	Weiner	Patrick	Feinstein	Palin	Pelosi
Barack Obama	3,675,192									
Mitt Romney	73,280	943,005								
Howard Dean	20,160	924	25,764							
Joe Biden	219,955	4,487	8,095	230,554						
Mike Bloomberg	12,873	2,658	831	2,408	20,076					
Anthony Weiner	31,169	2,158	4,618	7,437	1,270	42,211				
Deval Patrick	17,523	1,076	1,619	3,751	562	1,164	21,608			
Diane Feinstein	8,590	484	1,359	2,727	308	1,084	518	11,589		
Sarah Palin	142,022	627,793	1,229	6,356	2,748	3,066	1,006	634	1,649,936	
Nancy Pelosi	33,467	2,562	7,280	10,185	1,196	6,117	1,660	2,363	3,429	41,690

Note: Diagonal entries are the number of fans of the page. Off-diagonal entries are the number of fans of both pages.

**TABLE 3. The First Ten Rows of the Ratio of Affiliation Matrix**

	Obama	Romney	Dean	Biden	Bloomberg	Weiner	Patrick	Feinstein	Palin	Pelosi
Barack Obama	1.000	0.020	0.005	0.060	0.004	0.008	0.005	0.002	0.039	0.009
Mitt Romney	0.078	1.000	0.001	0.005	0.003	0.002	0.001	0.001	0.666	0.003
Howard Dean	0.782	0.036	1.000	0.314	0.032	0.179	0.063	0.053	0.048	0.283
Joe Biden	0.954	0.019	0.035	1.000	0.010	0.032	0.016	0.012	0.028	0.044
Mike Bloomberg	0.641	0.132	0.041	0.120	1.000	0.063	0.028	0.015	0.137	0.060
Anthony Weiner	0.738	0.051	0.109	0.176	0.030	1.000	0.028	0.026	0.073	0.145
Deval Patrick	0.811	0.050	0.075	0.174	0.026	0.054	1.000	0.024	0.047	0.077
Diane Feinstein	0.741	0.042	0.117	0.235	0.027	0.094	0.045	1.000	0.055	0.204
Sarah Palin	0.086	0.380	0.001	0.004	0.002	0.002	0.001	0.000	1.000	0.002
Nancy Pelosi	0.803	0.061	0.175	0.244	0.029	0.147	0.040	0.057	0.082	1.000

Note: Because each page has a different number of fans, the denominator changes and the off diagonal entries are not identical.

The maximum number of fans of a page is 3.67 million (Barack Obama), with an average of 15,422.5 fans per page. Most pages have a few thousand fans.<sup>6</sup>

To estimate ideology our approach is similar to the approach used by Aleman et al. (2009) to estimate the ideal points of legislators in the United States and Argentina using cosponsorship data, and is similar in principle to the use of Relational Class Analysis by Goldberg (2011) and Baldassarri and Goldberg (Forthcoming). Estimating separate parameters for users and pages, as is typical of estimation techniques like DW-NOMINATE or Bayesian analysis (Clinton, Jackman, and Rivers 2004), would be difficult on a large, sparse matrix. Instead, we construct an affiliation matrix between the political pages in which each cell indicates the number of users that like both pages. We do not use the original (two-mode) dataset of connections between users and political figures, which is organized as an  $X = r \times c$  matrix, with  $r = 1, 2 \dots R$  users and  $c = 1, 2 \dots C$  pages, but instead use an affiliation matrix (Table 2),  $A = XX'$ . In this affiliation matrix, the diagonal entries are the total number of users that like each page and the off-diagonal entries are the number of times an individual user likes both pages. Table 2

shows the first ten rows and ten columns of the affiliation matrix,  $A$ . The table shows that there are very significant differences in the total number of users that like each page, as well as notable differences in the number of users that like each pair of pages.

Next, we calculate the ratio of shared users by dividing the number of users that like *both* pages by the total number of users that like *each* page independently, which produces an agreement matrix,  $G = a_{ij} / \text{diag}(a_i)$ , as depicted in Table 3. Because each page has a different number of fans, the denominator changes and the upper and lower triangles of the new square matrix,  $G$ , are not identical. For instance, Barack Obama has many fans on Facebook (the most of any page in the data set) so the values in the first row are small as they are all divided by the number of fans that Obama has. However, many of the fans of other candidates also like Obama, so those values are relatively high.

It is notable that there is overlap in fans across partisan lines. Take, for instance, the Barack Obama column in Table 3: this column represents the proportion of the other politicians' fans who are also Obama's fans. Although it is not surprising that the other Democratic politicians have fans that are also Obama's fans, more than 7% of both Romney's and Palin's fans are also fans of Obama. Thus, for Facebook users who are fans of at least two candidates, there is not complete

<sup>6</sup> These figures are current as of March 2012.

polarization among Facebook users. This may owe to centrist users who are interested in following members of both parties or to users who are interested enough in politics to follow a large set of popular politicians regardless of their ideological affiliation.

The agreement matrix provides the information required for estimation. From this stage, a number of methods to scale the data may be employed. For simplicity, we use SVD on the centered matrix,  $G$ . Because we normalize the agreement matrix, which makes it asymmetric, the results from the left and right singular vectors are not similar. The right singular value is still highly related to the page's popularity, as its denominator is the number of fans of the page. The left singular value is unrelated to popularity, as its denominator is unrelated to the popularity of the page. Therefore, we retrieved the first rotated left singular value as the ideology measure for the pages. We rescaled the values to have mean 0 and standard deviation 1 for ease of interpretation.

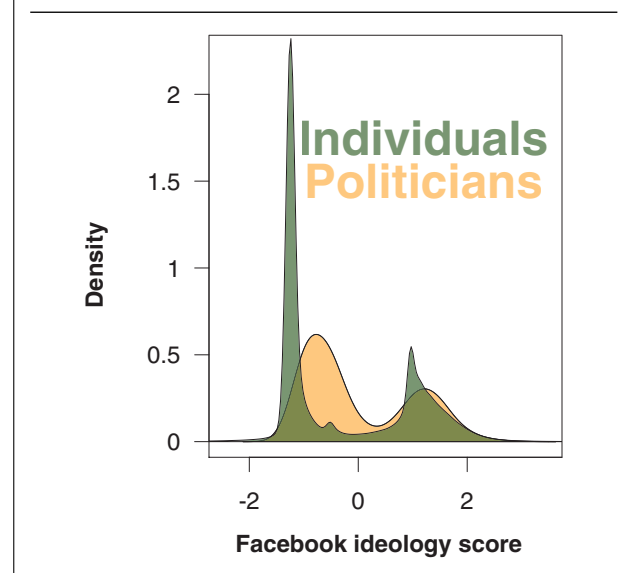
We were next interested in estimating ideology scores for the users. If we were to scale the entire user by page matrix we would be able to estimate separate parameters for the users and the pages that should be measures of ideology. Another approach would be to replicate what we have done with the pages and to create a matrix of connections between users based on shared political pages that they both like. This would create a matrix of approximately  $6.2 \text{ M}^2$  entries. However, decomposing a  $6.2\text{-M} \times 6.2\text{-M}$  matrix generally requires more computational resources than are available to political scientists running R on a normal computer. Instead, for each user we take the average of the scores for the political pages that user endorsed.

We begin our exploration of our estimates by examining the distributions of the ideology scores of politicians and individuals, as shown in Figure 2. The figure shows that both individuals and politicians are bimodally distributed, which is similar to Poole and Rosenthal's (1997) results for the U.S. Congress, and Bonica's (2013) results for candidates for office. The bimodal distribution of individuals is consistent with a polarized American public (Abramowitz 2010; Abramowitz and Saunders 2008; Levendusky 2009), in contrast to those who have argued that polarization in the electorate has not increased (Dimaggio, Evans, and Bryson 1996; Fiorina, Abrams, and Pope 2006; McCarty, Poole, and Rosenthal 2006). However, our sample is not fully representative of the population. Indeed, because those in our sample are likely to be more politically engaged than average, our data may reflect that more politically engaged citizens are also more polarized (Abramowitz and Jacobson 2006; Fiorina and Levendusky 2006). Finally, while the distributions of both sets of actors show evidence of polarization, politicians are more dispersed than individuals.

## VALIDATION OF THE MEASURE

In order to validate the measures that we estimate, we begin by showing the correlation between our mea-

**FIGURE 2. Density Plots of Ideological Estimates of 1,223 Politicians and 6.2 million Individuals.**



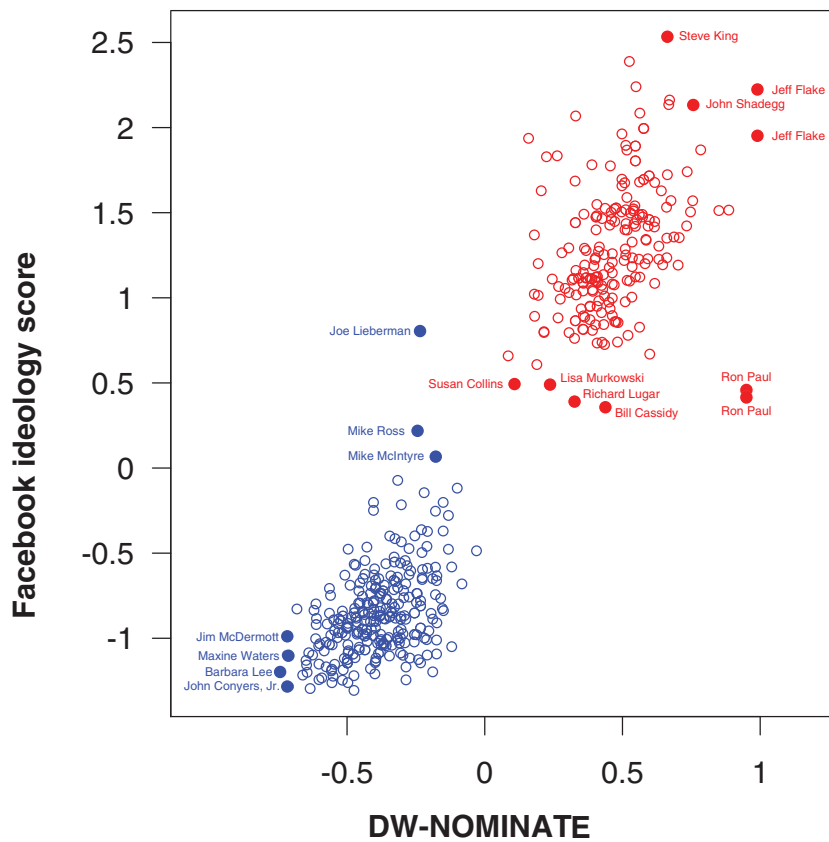
asures and other commonly used measures. For political elites, we rely on measures of ideology that come from their voting records. For individuals, we use self-reported indicators of ideology.

## Validation of Measures for Political Elites

To examine the similarity between legislators' positions as derived from roll-call vote data and those from Facebook liking data, we matched Facebook pages to their corresponding DW-NOMINATE first dimension scores. We matched 465 pages to members of the 111th Congress. The Pearson correlation between the two measures is 0.94, and the within-party correlation for Democrats is 0.47 and for Republicans is 0.42. This correlation is quite high given that Aleman et al. (2009) find that the ideological estimates from roll-call data and ideological estimates from cosponsorship data in the U.S. Congress correlate between 0.85 and 0.94. Figure 3 provides a visual representation of the relationship between the two ideology scores. The figure shows that the measures cluster legislators into two parties, and that the correlation within parties is quite high as well. For comparison, Bonica (2013) finds that the overall correlation between DW-NOMINATE and CFscores scores among incumbents is 0.89, with a Democratic correlation of 0.62 and a Republican correlation of 0.53.

We have labeled some of the members of Congress in the figure to illustrate where some of the most extreme members and some of the more moderate members lie on both measures. There are two points for Jeff Flake and Ron Paul in Figure 3. Although most of the political figures in the dataset had only one page, some had more than one. Ron Paul maintained two official pages, one for his presidential candidacy and one for

**FIGURE 3. Scatter Plot Showing the Relationship between the Facebook Based Ideology Measure and DW-NOMINATE**



his work in Congress (this page explicitly asked for all discussion related to his presidential candidacy to move to the other page). Ethics rules state that members need to separate official business and campaign activity, and these separate pages may reflect an effort to comply. While it is unfortunate to have multiple pages representing the same individual with respect to reliably estimating ideology, it does allow us to see whether we get consistent ideological estimates across multiple pages. The similarity in ideological estimates for Flake and Paul’s pages suggests that our measurement strategy is reliable.

**Validation of Measures for Ordinary Citizens**

We next turn to the validation of the individual-level ideological estimates. First, we computed the average ideology score of individuals based on their stated political views. On Facebook users may fill out a free response field that many users fill out called “political views.” Many users type the same things in, such as “Democrat,” “Republican,” “Liberal,” or “Conservative.” We took all labels that more than 20,000 users had used and calculated the average ideology score for

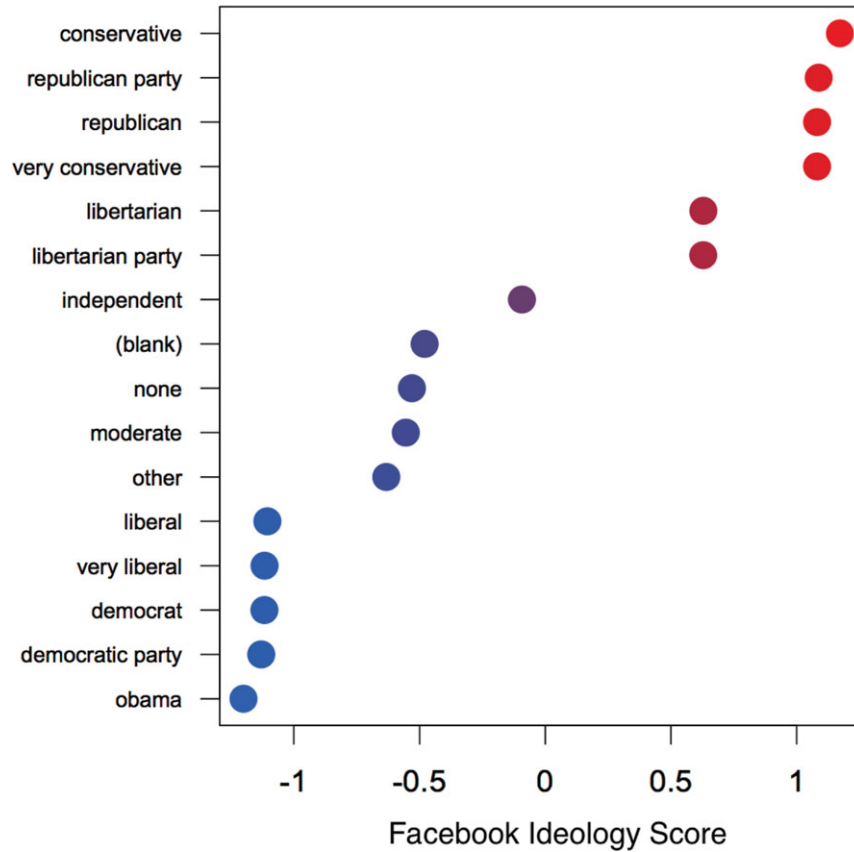
the group, as well as the 95% confidence interval for that estimate. The results are shown in Figure 4.

Figure 4 shows that the ideology score predicts users’ stated political views well. There appear to be at least three clear groups—those who state their political views and are liberal, those who do not state a clearly liberal or conservative ideology, and those who state their ideology and are conservative. There is also substantial variation in the middle group, those that do not state a liberal or conservative ideology. The groups represented are in approximately the order one would expect based on their average ideology, save for the fact that those who self-identify as “very conservative” are slightly to the left of those who self-identify as “conservative.”

While the above analysis suggests that the estimates we make for users are valid, stated political views on Facebook is a measure of political orientation that has not been previously analyzed. Stated political views of an individual on Facebook are difficult to interpret because we do not know how users understand the question. Therefore, we conducted a survey using a more standard ideology survey question to further understand if the estimates correlate with this commonly used ideology metric.



**FIGURE 4. Average Facebook Ideology Score of Users Grouped by the Users' Stated Political Views**



Notes: The category labeled “none” is the group of users that actually wrote the word “none” as their political views. The point labeled “(blank)” is the group of users that has not entered anything in as their political views. The 95% confidence intervals for each of the estimates is smaller than the point. The color of the points is on a scale from blue to red that is proportional to each group's average ideology score.

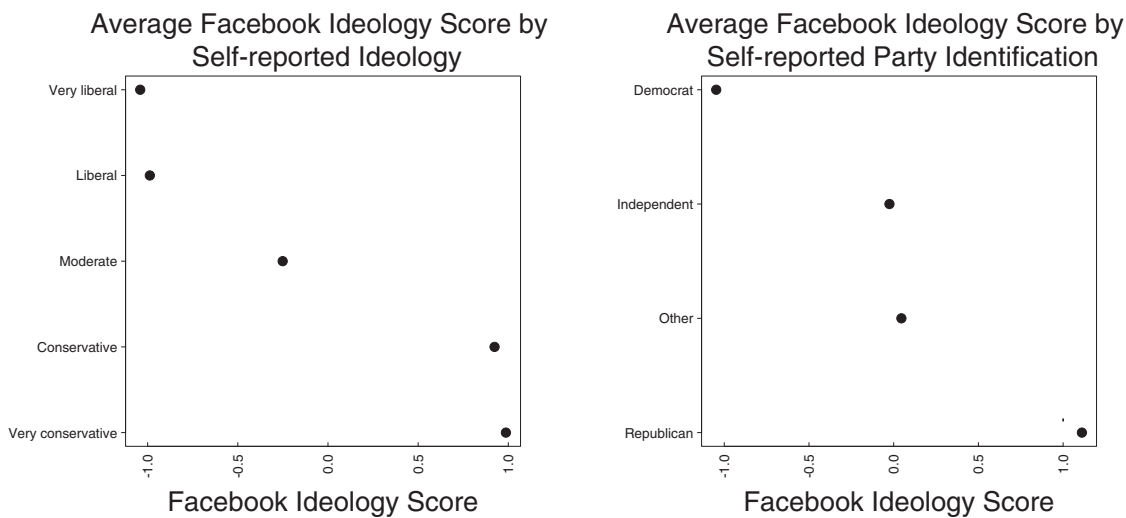
We use survey data from 20,027 individuals for whom we were also able to estimate ideology to further validate our ideology measure. These individuals consist of the intersection between individuals who took the survey and those who liked two or more pages. The survey was issued to a convenience sample on September 6, 2012 to U.S. Facebook users who were logged in. 282 thousand individuals started the survey and 78 thousand completed it; the approximate AAPOR response rate was 2.6 percent. Respondents' demographics reflect a departure from U.S. Census demographics with respect to age (54% 18–34, 34% 35–54, and 11% 55+); gender (57% female); ethnicity (83% white); education (19% high school or less; 41% some college; 28% college graduate; 12% postgraduate); and, to a lesser extent, ideology (12% very liberal; 22% liberal; 42% moderate; 17% conservative; 6% very conservative).

We asked individuals to place themselves on a five point ideological scale and also for their party identifi-

cation as either a Democrat, Republican, Independent, or Other. We then computed the average Facebook ideology score for each group. The results are shown in Figure 5.

Figure 5 shows that the Facebook ideology score is strongly associated with traditional survey ideology and partisanship measures. The results show there are three clear groups: Liberals, Moderates, and Conservatives. Although the measure can statistically differentiate between those who answered “Liberal” from those who answered “Very liberal” and those who answered “Conservative” from those who answered “Very conservative,” the differences between those groups is small. Further, the ideology measure from Facebook is a good predictor of partisanship, with Democrats occupying the scale's left end, Republicans the right, and Independents the middle. We do not expect the respondents to our survey to be representative of the population overall or of the population for which we measure ideology, but we also do not expect that either

**FIGURE 5. Average Facebook Ideology Score of Users Grouped by the Users' Ideology (left panel) and Party Identification (right panel) from a Survey Conducted Through the Facebook Website**



Note: The 95% confidence intervals for each of the estimates is smaller than the point.

method will show bias in *measurement* for either population. Therefore, we use these data to validate the Facebook measure, not to make arguments about the overall population.

## EXAMPLE APPLICATIONS OF THE IDEOLOGY MEASURE

### Example 1: Age and Ideology

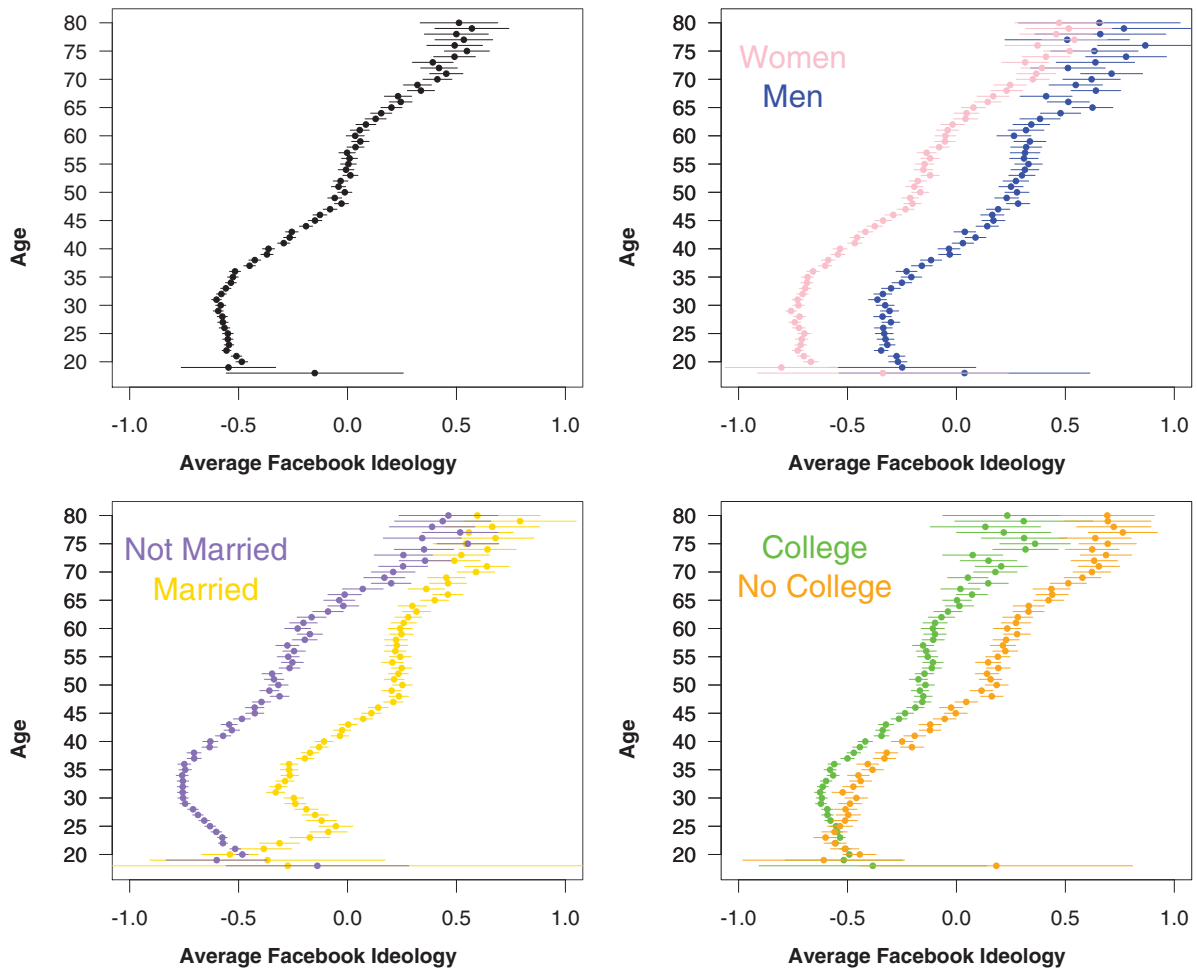
A substantial literature has found a relationship between age and political ideology (Cornelis et al. 2009; Glenn 1974; Krosnick and Alwin 1989; Ray 1985). Most studies have found that as people age they become more conservative and less susceptible to attitude change (Jennings and Niemi 1978; Krosnick and Alwin 1989). However, less is known about how that relationship varies across individual characteristics. Recent studies have begun to investigate how personality plays a role in mediating the relationship between age and conservatism (Cornelis et al. 2009). Here we investigate how the relationship between age and ideology varies by characteristics such as gender, marital status, and educational attainment.

A key advantage of using the Facebook ideology data is the large number of observations. By looking at patterns in the raw data we can understand phenomena that are not as easily understood using standard techniques, such as regression. In order to study ideology across characteristics, we took all individuals for whom we calculated an ideology score and matched the individual's characteristics that they listed on their

profile. Figure 6 shows the average ideology of users age 18 through 80, then separates the estimates by gender, marital status, and college attendance. The figure shows that older people are more conservative than younger people. Women are more liberal than men but a similar pattern in which older women and older men are more conservative than their younger counterparts emerges. While the overall pattern is similar for those who get married and those who do not, young people who get married are more conservative than their younger counterparts and young people who do not get married are more liberal than their younger counterparts. After the age of approximately 35, the pattern of increasing conservatism with age is similar across the groups. Finally, college attendance is not predictive of ideology among the young, but for among older groups having attended college is related to being more liberal.

While these results are consistent with previous research, we have not studied change in ideology over time. The finding that older people are more conservative than younger people is consistent with a population that becomes more conservative over time. However, it is also consistent with recent surveys that have shown that the most recent generation is among the most liberal in recent memory (Kohut et al. 2007). While we have some evidence of how an individual's ideology changes over time, we find that the correlation in ideology from March 2011 to March 2012 is 0.99. With more time we should be able to better discern whether the patterns we found are due to cohort effects or change in an individual's ideology over time.

**FIGURE 6. In Each Panel the Points Show the Average Ideology of the Age Group for Individuals Age 18 through 80, and the Lines Represent the 95% Confidence Interval of the Estimate**



Notes: The upper left panel shows the average ideology of all users in our sample. The upper right panel shows the average ideology of men and women by age. The lower left panel shows the average ideology of married and unmarried individuals by age. The lower right panel shows the average ideology of college attendees and those who have not attended college by age.

### Example 2: Ideology in Social Networks

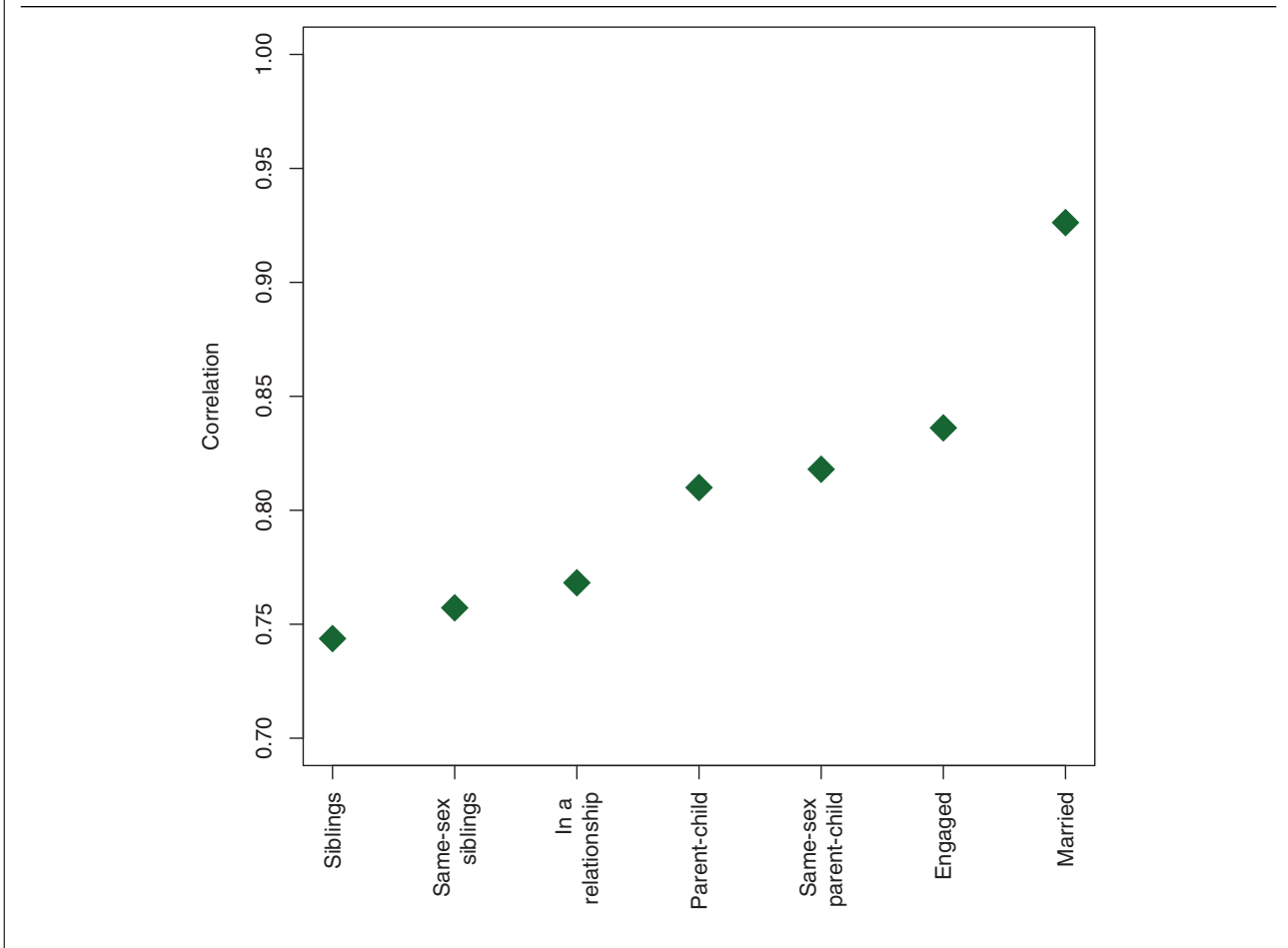
Another advantage of using the Facebook ideology data is the abundance of data about the social networks of its users. Previous work has shown that ideology clusters in social networks (Huckfeldt, Johnson, and Sprague 2004). Here, we wish to characterize the extent to which the clustering varies based on the strength of the relationship between two individuals. We consider three general types of relationships: friendships, family ties, and romantic partners.

Clustering in the network may be due to some combination of three possibilities. First, clustering may be due to exposure to a shared environment. Friends may be both exposed to some external factor that influences both to change their ideology in a similar way (e.g., attending the same college (Newcomb 1943)). Second, clustering may be due to homophily. That is, people may choose friends based on ideology (Heider

1944; Festinger 1957). Third, clustering may be due to influence. That is, one friend may argue in favor of an ideological position and change their friend's views (Lazer et al. 2010). These processes are more likely to occur between close friends than socially distant ones—closer friends are more likely to be physically proximate and hence more likely to experience the same external stimuli. Friends who share an ideology are more likely to become close friends as they will have more in common. Finally, close friends will have more opportunities to influence one another's ideological views, which should also increase ideological similarity.

As social relationships become stronger, friends should be increasingly likely to hold similar ideological views. And, this pattern should be strongest for the most intimate relationships. Recent work has shown that while people do not usually select romantic partners based on ideology, they do base such

**FIGURE 7. The Correlation in Ideology for Familial and Romantic Relationships. The 95% Confidence Intervals for Each of the Estimates is Smaller than the Point**



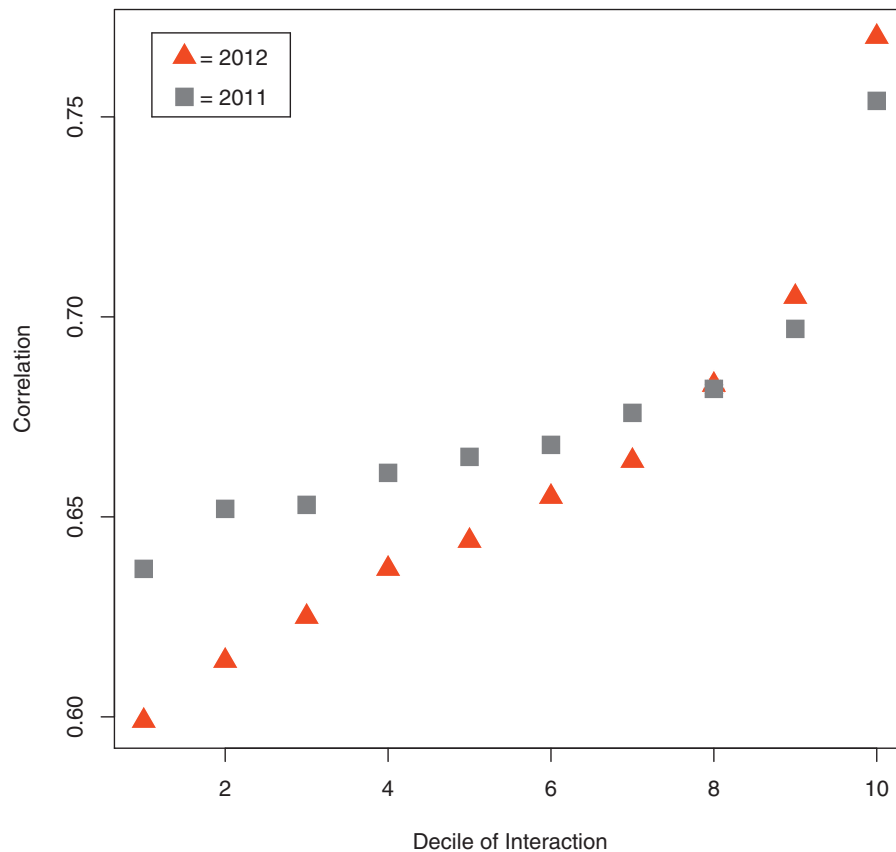
decisions on factors related to ideology, which leads to highly correlated ideological views (Klofstad, McDermott, and Hatemi 2013). Furthermore, this pattern should grow stronger over time as partners are exposed to the same factors, influence each other, and as more similar partners should be more likely to stay together.

Similarly, we expect that familial relationships will show evidence of these processes. Members of a nuclear family are more likely to experience the same external stimuli, due to being more likely to live near one another. While parents cannot select their children (or vice versa) based on ideological views (or other characteristics related to it), recent work shows a genetic component to ideology (Hatemi et al. 2011). This genetic component coupled with socialization and similar exposure to factors that influence ideology mean that ideology is likely to be correlated within family units.

Since Facebook allows users to identify their familial and romantic relationships, we were able to test for ideological similarity across these social links. We began by pairing all individuals with their siblings, parents, or romantic partners. We then calculated the Pearson

correlation for each group. The results are shown in Figure 7. The figure shows that married couples have the highest correlation in ideology, while engaged couples have the second highest value. Correlations within the nuclear family have lower values, with parent-child relationships being stronger than sibling relationships. The fact that the correlation between siblings is lower than the correlation between parent-child pairs may be due to the fact that our sample skews toward younger users and that the set of parents on Facebook are more likely to be similar to their children across a range of factors than parents who have not yet joined the website.

Next, we were interested in the correlation of ideology between friends. First, we paired all 6.2 million users for whom we estimated ideology in 2012 with every Facebook friend for which we also had an estimate for ideology, for a total of 327 million friendship dyads and an average of 53 friends per user. The overall correlation in 2012 was 0.69, which approximates other measures of ideological correlation among friends (Huckfeldt, Johnson, and Sprague 2004). We repeated this procedure for 2011, with 6.1 million users who had 238 million friendship connections to other users for

**FIGURE 8. The Correlation in Ideology for Friendship Relationships**

Notes: Each decile represents a separate set of friendship dyads. Decile of interaction is based on the proportion of interaction between the pair during the three months before ideology was scaled. The 95% confidence intervals for each of the estimates is smaller than the point.

whom we estimate ideology.<sup>7</sup> The overall correlation between ideology among these friends in 2011 was 0.67. The slight increase in correlation of friends' ideology from 2011 to 2012 is suggestive that there is greater polarization among friends in 2012.

Additionally, we categorized all friendships in each year of our sample by decile, ranking them from lowest to highest percent of interactions. Each decile is a separate sample of friendship dyads. We validated this measure of tie strength with a survey (see Jones, Settle, Bond, Fariss, Marlow, and Fowler (2012) for more detail) in which we asked Facebook users to identify their closest friends (either 1, 3, 5, or 10). We then measured the percentile of interaction between friends in the same way and predicted survey response

based on interaction between Facebook friends. The results show that as the decile of interaction increases the probability that a friendship is the user's closest friend increases. This finding is consistent with the hypothesis that the closer a social tie between two people, the more frequently they will interact, regardless of medium. In this case, frequency of Facebook interaction is a good predictor of being named a close friend.

Using the decile measure of tie strength, we then calculated the correlation between user and friend ideology on each set of dyads for both 2011 and 2012 (see Figure 8). For both years the correlation in friends' ideology increases as tie strength increases. The proportion of interaction between friends is a better predictor of similarity of ideology between friends in 2012 than in 2011, again suggesting that in 2012 friendships are more politically polarized than in 2011. We caution that this correlation may increase for other reasons, such as better measures of ideology in 2012 owing to users liking more political pages, or perhaps to changes of the makeup of the set of individuals for whom we can estimate ideology. Although we can only estimate ideology for about 2% more people in 2012 than 2011,

<sup>7</sup> The correlation over time of ideology within an individual from 2011 to 2012 is 0.99. While this is further evidence of ideological constraint, it is important to consider the measure's construction, as it can impact the measure of change over time (Achen 1975; Converse 2006). As few users changed the politicians they liked over the period, such a high correlation in ideology not entirely surprising. The correlation is not due only to few users changing the pages they like, but also the changes that did occur did not change the ordering of the ideologies of the pages to a significant degree.

the change in the sample could be enough to account for the difference in 2012.

### Example 3: Friend Ideology and Turnout

While the composition of ideology in social networks is important on its own, considerable research has been applied to understanding how the makeup of an individual's social network affects political participation (Eveland and Hively 2009; Huckfeldt, Johnson, and Sprague 2004; Huckfeldt, Mendez, and Osborn 2004; McClurg 2006; Mutz 2002; Scheufele et al. 2004). Scholars have long theorized about how cross-cutting pressures in an individual's social environment may cause an individual to become less interested in politics and to disengage (Campbell et al. 1960; Ithel de Sola Pool and Popkin 1956). Recent work has tested theories about whether disagreement in an individual's social network affects the political participation (Mutz 2002; Huckfeldt, Johnson, and Sprague 2004). This work has consistently found that exposure to disagreement depresses engagement and participation.

Previous studies have relied on snowball samples in order to construct social network measures. Survey respondents may be biased in recalling their discussion partners. Social network sites, such as Facebook and Twitter, allow us to observe friendships without asking individuals about their political discussion partners. If bias in recalling discussion partners is associated with the difference in ideology between a pair of individuals, which certainly may be the case if such discussions are more likely to be memorable, estimates of its effect on participation may be biased as well.

We add to this literature by studying the relationship between exposure to disagreement and validated public voting records. We matched records for all individuals from 13 states (see Jones, Bond, Fariss, Settle, Kramer, Marlow, and Fowler (2012) for more information) to the Facebook data. Among the 6.2 million individuals for whom we calculated an ideology score, we match a voting record for 397,815,<sup>8</sup> resulting in a total of 2,410,097 friendship pairs where the individual and the friend had both an ideology score and a public record of whether each person voted. Here we test the relationship between an individual's (in social network terms, the "ego") turnout and the difference between the ego's ideology and that of a friend (the "alter"). Our primary variable of interest is the *Mean |Ideology Difference|*, as it measures the average difference in ideology between the ego and all connected alters. As Table 4 shows, an increase in the ideological distance between friends is associated with lower rates of turnout by the ego. A one standard deviation increase in the *Mean |Ideology Difference|* measure corresponds to a 1.3 [95% CI 1.1, 1.4] percentage point decrease in the likelihood of voting (95% CI estimated based on King, Tomz, and Wittenberg 2000).

<sup>8</sup> The low match rate of about 6% owes to the fact that we only have voting information from 13 states, while our sample for which we calculate ideology scores may reside in any state.

**TABLE 4. Logistic Regression of Ego Validated Voting in 2010 on Ego Covariates and Alter Characteristics of Ideology and Turnout.**

(Intercept)	− 2.325*** (0.210)
Mean  Ideology Difference	− 0.085** (0.028)
Ego Age	0.065*** (0.009)
Ego Age Squared	− 0.001*** (0.001)
Prop. Alter Voted	1.016*** (0.062)
Ego Ideology Est.	0.199*** (0.022)
Ego Ideology	0.426*** (0.071)
Ego Female	− 0.058 (0.045)
Ego Married	0.400*** (0.043)
Ego College	0.399*** (0.047)
Ego Friend Count	0.009 (0.016)
Nagelkerke R-sq.	0.162
Log-likelihood	−206 266.566
Deviance	412 533.133
AIC	412 555.133
BIC	412 674.456
N	379 876

Note: Robust standard errors clustered by common friend (e.g., Williams 2000; Wooldridge 2002).

This result is consistent with previous work showing that disagreement in an individual's social network is associated with lower turnout rates (Huckfeldt, Johnson, and Sprague 2004; Mutz 2002). However, the present work has the advantage of gathering behavioral data. That is, we use validated public voting records, a behavioral ideology measure, and observe friendships, which help us avoid an array of issues including survey response bias and recall bias for friendships.

## DISCUSSION

This article makes several important contributions: first it presents a method for measuring ideology using large-scale data from social media, and second, it uses that measure to examine political polarization, the structure of ideology that characterize relationships in society, and how that structure is associated with participation rates. We show that the method produces reliable ideological estimates that are predictive of other measures of elite ideology—DW-NOMINATE, and individual level ideology—self-reported liberal-conservative ideological background. Placement of elites and masses on the same scale is an important step in the study of electoral politics and political communication. For instance, a longstanding debate in the literature concerns whether ordinary

citizens are polarizing (Abramowitz and Saunders 2008; Fiorina, Abrams, and Pope 2006) and, if so, whether this divide is driven by elites or mass preferences. Data that put elites and ordinary citizens on the same scale are necessary to the study of these types of phenomena because they allow for reliable comparison of the ideology of both groups.

We then investigate the ideological polarization of social relationships by examining how ideology maps onto the structure of our social relationships, coupling our ideological measure with extensive data about social ties and interaction online. We not only confirm previous work finding that friendships cluster by ideological preference, but extend this work to show that ideological correlation is stronger between close friends, family, and especially among romantic partners. Further, we show that from 2011 to 2012 there is an increase in the extent to which ideology is associated with the closeness of a friendship, suggesting polarization over the one-year period. This evidence contributes to the debate about polarization by using new evidence about not only the distribution of ideologies in the electorate, but also about the distribution of social relationships among those with similar and different ideological preferences. A better understanding of this type of polarization is critical, as the ideologies of our social contacts can impact the likelihood that we are exposed to new ideas, which is a critical component of democracy (Huckfeldt, Johnson, and Sprague 2004). Future work on polarization among friends will be important for understanding the evolution of ideological polarization in ordinary citizens.

Finally, we study the association between disagreement in an individual's social network and decreased rates of turnout. This result is consistent with previous work (Huckfeldt, Johnson, and Sprague 2004; Mutz 2002), but has the advantage of using behavioral measures of ideology, turnout, and friendships. This helps to avoid biases that individuals have in answering survey questions about ideology and turnout and recalling friends with whom they have discussed politics. While our results show that individuals in networks with disagreement are less likely to participate in politics, further work should further investigate whether this relationship is causal.

It is important not to lose sight of the potential drawbacks of the approach we use and the weaknesses of this particular study. First, the population we study here is not representative of the U.S. population overall. For instance, Facebook users are on average younger than the population. However, more than 100 million American adults are Facebook users, out of a population of 230 million. Furthermore, the particular sample we use here—those who have liked two or more political pages—differs from the overall population of Facebook users. This subset is likely older, more politically engaged, and more engaged with Facebook than the common user. However, our sample is not limited to political elites or donors to political causes. Rather, they constitute a subset of the population who are on average simultaneously more interested in politics and more engaged on Facebook.

Early in the article, we discussed several problems with traditional measures of individual ideology. First, traditional survey measures do not account for the multiple dimensions of ideology. While we have focused here on the first dimension of ideology, our method of measurement recovers many dimensions related to ideology. In future work, scholars should investigate these dimensions and how they can help us to better understand ideology outside of the traditional left-right paradigm—for example, can this measure help us understand libertarian-authoritarian and “intervention-free market” dimensions (Hix 1999)? Second, traditional survey measures rely on an interval ideology measure, which often assumes that differences between responses are equal (that is, the difference between a response of “Very Liberal” and “Liberal” is assumed to be the same as the difference between responses of “Liberal” and “Somewhat Liberal”). While our measure is continuous by its construction, we find that there are three clustered groupings of individuals: those who are liberal, those who are conservative, and those who are somewhere in the middle. While we find few differences between those who label themselves “Liberal” or “Very Liberal” and “Conservative” or “Very Conservative,” future work should further investigate whether the lack of meaningful difference is due to our measurement strategy or to other factors related to how individuals present themselves politically online.

Finally, traditional survey measures of ideology are unreliable, particularly among liberals who are unwilling to label themselves as such. The construction of our measure should help to make the measure more reliable. Still, we cannot say that our measure is strictly better than others. Where survey respondents may give misleading answers to questions about ideology, in our case, individuals may like political pages to acquire information about a candidate in addition to signaling support. As scholars investigate how people make decisions about how to present themselves politically online we will be better able to understand how any issues affect the reliability of our measure.

Our hope is that the method we have described here will allow for tests of theories from spatial models of politics that require data on the relative positioning of political actors, as have previous methods that put a diverse set of political actors on a common scale. Our method has the potential to uncover ideological estimates from any entity present in social media—including legislators, the candidates for office they have defeated, bureaucrats, ballot measures, and political issues. This type of data will be important for our ability to study political phenomena concerning the interaction between legislator and constituent ideology, such as representation or vote choice.

The possibilities for future research on large data sets that contain previously unstudied types of information about people such as Facebook and Twitter should not be underestimated. The increased power associated with having a large number of individuals affords researchers the opportunity to unobtrusively test theories we previously could not; furthermore it

increases precision when we do so.<sup>9</sup> While in many cases, such fine-grained estimates may not be necessary and conventional survey sampling works well, there are an array of applications for which such estimates are desirable: estimates for ideology or support for candidates in small districts, local officials; estimates of support among minority populations; estimates of change over short periods of time; and ideological estimates in non-U.S. political locales where reliable survey, roll-call, and/or fund-raising data might not be available. With measures such as the one described in this article, we may be able to detect changes in the public perception of political officials and/or candidates simply by examining the public's changing preferences online.

This research is part of a growing literature in the social sciences in which large sources of data are used to conduct research that was previously not possible (Lazer et al. 2009). We hope that the ideology measure we use in this article, and others that measure the ideology of large numbers of the mass public (Bonica 2013; Tausanovitch and Warshaw 2012), will contribute to our understanding of ideology in new ways. While there has been a long tradition of research into ideology and its structure, this article should form a starting point for future research into how our social networks are critical to our understanding of society's ideological makeup.

## NOTE

Due to privacy considerations, data must be kept on site at Facebook. However, Facebook is willing to give access to de-identified data on site to researchers wishing to replicate these results.

## REFERENCES

- Abramowitz, Alan. 2010. *The Disappearing Center: Engaged Citizens, Polarization and American Democracy*. New Haven: Yale University Press.
- Abramowitz, Alan, and Gary C. Jacobson. 2006. *Red and Blue Nation? Characteristics and Causes of America's Polarized Politics*. Washington, DC: Brookings Institution Press, chapter "Disconnected or Joined at the Hip?"
- Abramowitz, Alan, and Kyle Saunders. 2008. "Is Polarization a Myth?" *Journal of Politics* 70 (2): 542–55.
- Achen, Christopher. 1975. "Mass Political Attitudes and the Survey Response." *American Political Science Review* 69 (4): 1218–31.
- Aleman, Eduardo, Ernesto Calvo, Mark P. Jones, and Noah Kaplan. 2009. "Comparing Cosponsorship and Roll-Call Ideal Points." *Legislative Studies Quarterly* 34 (1): 87–116.
- Ansolabehere, Stephen, Jonathan Rodden, and James M. Snyder. 2008. "The Strength of Issues: Using Multiple Measures to Gauge Preference Stability, Ideological Constraint, and Issue Voting." *American Political Science Review* 102 (2): 215–32.
- Bafumi, Joseph, and Michael Herron. 2010. "Leapfrog Representation and Extremism: A Study of American Voters and their Members in Congress." *The American Political Science Review* 104: 519–42.
- Bailey, Michael. 2007. "Comparable Preference Estimates Across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51: 433–48.
- Baldassarri, Delia, and Amir Goldberg. 2014. "Neither Ideologues, nor Agnostics: Alternative Voters' Belief System in an Age of Partisan Politics." *American Journal of Sociology* (forthcoming).
- Barberá, Pablo. 2014. "Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data." Political Analysis (forthcoming).
- Bond, Robert M., Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. 2012. "A 61-million-person Experiment in Social Influence and Political Mobilization." *Nature* 489 (7415): 295–8.
- Bonica, Adam. 2013. "Ideology and Interests in the Political Marketplace." *American Journal of Political Science* 57 (2): 294–311.
- Butler, Daniel M., and David E. Broockman. 2011. "Do Politicians Racially Discriminate Against Constituents? A Field Experiment on State Legislators." *American Journal of Political Science* 55 (3): 463–77.
- Butler, Daniel M., Christopher F. Karpowitz, and Jeremy C. Pope. 2012. "A Field Experiment on Legislators' Home Styles: Service versus Policy." *The Journal of Politics* 74 (02), 474–86.
- Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. Chicago and London: University of Chicago Press.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *The American Political Science Review* 98 (2): 355–70.
- Converse, Philip. 2006. "The Nature of Belief Systems in Mass Publics." *Critical Review* 18 (1–3): 1–74.
- Cornelis, Ilse, Alain Van Hiel, Arne Roets, and Malgorzata Kosowska. 2009. "Age Differences in Conservatism: Evidence on the Mediating Effects of Personality and Cognitive Style." *Journal of Personality* 77 (1): 51–88.
- de Sola Pool, Ithel, Robert P. Abelson, and Samuel Popkin. 1956. *Candidates, Issues and Strategies*. Cambridge, MA: MIT Press.
- Dimaggio, Paul, John Evans, and Bethany Bryson. 1996. "Have Americans' Social Attitudes Become More Polarized?" *American Journal of Sociology* 102: 690–755.
- Eckart, Carl, and Gale Young. 1936. "The Approximation of One Matrix by Another of Lower Rank." *Psychometrika* 1 (3): 211–8.
- Eveland, Jr., William P., and Myiah Hutchens Hively. 2009. "Political Discussion Frequency, Network Size, and "Heterogeneity" of Discussion as Predictors of Political Knowledge and Participation." *Journal of Communication* 59 (2): 204–24.
- Facebook. 2011. "Statistics." <http://www.facebook.com/press/info.php?statistics>
- Feldman, Stanley. 1998. "Structure and Consistency in Public Opinion: The Role of Core Beliefs and Values." *American Journal of Political Science* 32 (2): 416–40.
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row Peterson.
- Fiorina, Morris, Samuel Abrams, and Jeremy Pope. 2006. *Culture War? The Myth of a Polarized America*. New York: Pearson Longman.
- Fiorina, Morris, and Matthew Levendusky. 2006. *Red and Blue Nation? Characteristics and Causes of America's Polarized Politics*. Washington, DC: Brookings Institution Press, chapter "Disconnected: The Political Class versus the People."
- Gerber, Elisabeth R., and Jeffrey B. Lewis. 2004. "Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation." *Journal of Political Economy* 112 (6): 1364–83.
- Glenn, Norval D. 1974. "Aging and Conservatism." *The ANNALS of the American Academy of Political and Social Science* 415 (1): 176–86.
- Goldberg, Amir. 2011. "Mapping Shared Understandings Using Relational Class Analysis: The Case of the Cultural Omnivore Reexamined." *American Journal of Sociology* 116 (5): 1397–436.
- Grimmer, Justin, Solomon Messing, and Sean J. Westwood. 2012. "How Words and Money Cultivate a Personal Vote: The Effect of Legislator Credit Claiming on Constituent Credit Allocation." *American Political Science Review* 106 (04): 703–19.
- Hatemi, Peter, Nathan Gillespie, Landon Eaves, Brion Maher, Bradley Webb, Andrew Heath, Sarah Medland, David Smyth, Harry Beeby, Scott Gordon, Grant Montgomery, Ghu Zhu,

<sup>9</sup> We caution that with this increase in statistical power that we should be careful to not confuse statistical significance with practical significance.



- Enda Byrne, and Nicholas Martin. 2011. "A Genome-Wide Analysis of Liberal and Conservative Political Attitudes." *Journal of Politics* 73 (1): 1–15.
- Heider, F. 1944. "Social Perception and Phenomenal Causality." *Psychological Review* 51 (6): 358–74.
- Hix, Simon. 1999. "Dimensions and Alignments in European Union Politics: Cognitive Constraints and Partisan Responses." *European Journal of Political Research* 35 (1): 69–106.
- Hix, Simon, Abdul Noury, and Gerard Roland. 2006. "Dimensions of Politics in the European Parliament." *American Journal of Political Science* 50 (2): 494–520.
- Huckfeldt, Robert, Paul E. Johnson, and John Sprague. 2004. *Political Disagreement: The Survival of Diverse Opinions within Communication Networks*. New York: Cambridge University Press.
- Huckfeldt, Robert, Jeanette M. Mendez, and Tracy L. Osborn. 2004. "Disagreement, Ambivalence, and Engagement: The Political Consequences of Heterogeneous Networks." *Political Psychology* 25 (1): 65–95.
- Jackson, John E. 1983. "The Systematic Beliefs of the Mass Public: Estimating Policy Preferences with Survey Data." *Journal of Politics* 45 (4): 840–65.
- Jennings, M. Kent, and Richard G. Niemi. 1978. "The Persistence of Political Orientations: An Over-Time Analysis of Two Generations." *British Journal of Political Science* 8 (3): 333–63.
- Jones, Jason J., Robert M. Bond, Christopher J. Fariss, Jaime E. Settle, Adam D. I. Kramer, Cameron Marlow, and James H. Fowler. 2012. "Yahtzee: An Anonymized Group Level Matching Procedure." *PLoS ONE* 8 (2): 55760.
- Jones, Jason J., Jaime E. Settle, Robert M. Bond, Christopher J. Fariss, Cameron Marlow, and James H. Fowler. 2012. "Inferring Tie Strength from Online Directed Behavior." *PLoS ONE* 8 (1): e52168.
- Kinder, Donald. 1983. *Diversity and Complexity in American Public Opinion*. Washington, DC: APSA Press, 389–425.
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 347–361.
- Klofstad, Casey, Rose McDermott, and Peter Hatemi. 2013. "The Dating Preferences of Liberals and Conservatives." *Political Behavior* 35: 519–38.
- Kohut, Andrew, Kim Parker, Scott Keeter, Carroll Doherty, and Michael Dimock. 2007. "How Young People View Their Lives, Futures and Politics: A Portrait of 'Generation Next'." <http://www.people-press.org/files/legacy-pdf/300.pdf>
- Krosnick, Jon A., and Duane F. Alwin. 1989. "Aging and Susceptibility to Attitude Change." *Journal of Personality and Social Psychology* 57: 416–25.
- Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *The American Political Science Review* 97 (2): 311–32.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323: 721–23.
- Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. 2010. "The Coevolution of Networks and Political Attitudes." *Political Communication* 27 (3): 248–74. <http://www.tandfonline.com/doi/abs/10.1080/10584609.2010.500187>
- Levendusky, Matthew. 2009. *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. Chicago: The University of Chicago Press.
- Malbin, Michael J. 2009. "Small Donors, Large Donors and the Internet: The Case for Public Financing after Obama." Unpublished manuscript.
- Martin, Andrew D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10 (2): 134–53.
- McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. 2006. *Polarized America: The Dance of Ideology and Unequal Riches*. Cambridge, MA: MIT Press.
- McClurg, Scott. 2006. "The Electoral Relevance of Political Talk: Examining Disagreement and Expertise Effects in Social Networks on Political Participation." *American Journal of Political Science* 50 (3): 737–54.
- Messing, Solomon, and Sean Westwood. 2012. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation when Selecting Online News Media." *Communication Research*, 41 (8): 1042–63.
- Monroe, Burt L., Michael Colaresi, and Kevin Quinn. 2009. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 15 (4): 372–403.
- Monroe, Burt L., and Ko Maeda. 2004. *Rhetorical Ideal Point Estimation: Mapping Legislative Speech*. Palo Alto: Stanford University.
- Mutz, Diana. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46 (4): 838–55.
- Mutz, Diana C., and Lori Young. 2011. "Communication and Public Opinion Plus Ca Change?" *Public Opinion Quarterly* 75 (5): 1018–44.
- Newcomb, Theodore Mead. 1943. *Personality & Social Change; Attitude Formation in a Student Community*. New York: Dryden Press.
- Nielsen. 2011. "State of the Media: Social Media Report Q3." <http://www.nielsen.com/content/corporate/us/en/insights/reports-downloads/2011/social-media-report-q3.html>
- Pew. 2012. "Internet Gains Most as Campaign News Source but Cable TV Still Leads: Social Media Doubles, but Remains Limited." [http://www.journalism.org/commentary\\_backgrounder/social\\_media\\_doubles\\_remains\\_limited](http://www.journalism.org/commentary_backgrounder/social_media_doubles_remains_limited)
- Poole, Keith. 2005. *Spatial Models of Parliamentary Voting*. New York: Cambridge University Press.
- Poole, Keith, and Howard Rosenthal. 1997. *Ideology and Congress*. New Brunswick, NJ: Transaction Publishers.
- Ray, John J. 1985. "What Old People Believe: Age, Sex, and Conservatism." *Political Psychology* 6 (3): 525–28.
- Scheufele, Dietram A., Matthew C. Nisbet, Dominique Brossard, and Erik C. Nisbet. 2004. "Social Structure and Citizenship: Examining the Impacts of Social Setting, Network Heterogeneity, and Informational Variables on Political Participation." *Political Communication* 21 (3): 315–38.
- Schiffer, Adam J. 2000. "I'm Not That Liberal: Explaining Conservative Democratic Identification." *Political Behavior* 22 (4): 293–310.
- Talbert, Jeffery C., and Matthew Potoski. 2009. "Setting the Legislative Agenda: The Dimensional Structure of Bill Cosponsoring and Floor Voting." *Journal of Politics* 64 (3): 864–91.
- Tausanovitch, Chris, and Christopher Warshaw. 2012. "Representation in Congress, State Legislatures and Cities." Unpublished manuscript.
- Treier, Shawn, and D. Sunshine Hillygus. 2009. "The Nature of Political Ideology in the Contemporary Electorate." *Public Opinion Quarterly* 73 (4): 679–703.
- Williams, Rick L. 2000. "A Note on Robust Variance Estimation for Cluster-Correlated Data." *Biometrics* 56 (2): 645–6.
- Wojcieszak, Magdalena E., and Diana C. Mutz. 2009. "Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?" *Journal of Communication* 59 (1): 40–56.
- Wood, Thomas, and Eric Oliver. 2012. "Toward a More Reliable Implementation of Ideology in Measures of Public Opinion." *Public Opinion Quarterly* 76 (4): 636–62.
- Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Boston: The MIT Press.