# *Someone is Wrong on the Internet*: Deliberation as the Responsive Exchange of Interconnected Ideas

Nick Beauchamp*
Northeastern University

June 10, 2014

**Abstract**

This paper develops a new model of political deliberation using a new dataset of millions of online posts arguing politics. Interpersonal political arguments are often taken to be unstructured ideological posturing with little persuasive effect, but this may in part be due to the difficulty of modeling the free-form speech characteristic of argument. Instead, this paper begins with a new psychological model structured as a network of interconnected ideas; estimates this cognitive network from online discussions using a new text-analytic model that combines techniques from Latent Dirichlet Allocation and semantic networks; uses the inferred network to predict topics deployed in political exchanges; and demonstrates that these arguments have long-term persuasive effects. This model of networked concepts, along with the short- and long-term dynamics, reveals a rich deliberative structure of political opinion and argument in ways impossible to discern with existing low-dimensional models of ideology and speech.

*Email: n.beauchamp@neu.edu; web: nickbeauchamp.com.

# 1 Introduction

> "Majority rule, just as majority rule, is as foolish as its critics charge it with being... The means by which the majority comes to be a majority is the important thing... The essential need, in other words, is the improvement of the methods and conditions of debate, discussion, and persuasion."
>
> - John Dewey, *The Public and Its Problems* (1946)

The great appeal of deliberative democracy is that it offers an improvement upon the dominant democratic model of majority rule (Dewey, Habermas, Rawls). Unlike mere aggregation of pre-existing preferences, deliberation can boost the voice and protect the rights of minorities, and aid society in producing decisions that are more rational, reasonable, and informed than voting alone might generate. More radically, deliberation may lead individuals to views that more genuinely reflect their fundamental preferences than did their pre-deliberative opinions Rawls???. The challenge for deliberative theorists has been to define the ideal process of deliberation as something separate from the outcomes it purports to achieve. To this end, a menagerie of criteria have been proposed: reason, reflexivity, sincerity and inclusion (Habermas 1994); seeking and providing information, a heterogeneity of views, and responsiveness to others viewpoints (Wilhelm 2000); mutual respect, justifications via external information, and discussion of the common good (Steenbergen et al. 2003); reflection upon a greater number of considerations and suitable emotions (Mansbridge et al. 2006); or various mixes of all of these things (Mutz 2008).[1] Unsurprisingly, given these many high bars both for process and outcomes, achieving these goals is often taken to require carefully designed and overseen environments (Fishkin 1991, Fishkin, Luskin & Jowell 2000, Mansbridge et al. 2006, Farrar et al. 2006, Mutz 2008), ???ackerman with an expense and complexity unlikely to be widely implementable.

More fundamentally, even when the resources for a proper deliberative poll exist, it is often unclear how to objectively assess the quality of the deliberative process independent of the outcome, even when a more narrow criterion has been selected. The more easily measured criteria, such as affect, view-changing, or turn-taking, seem farthest from the core notion of deliberation,

[1] Each of these lists is not exhaustive for the respective authors; we have highlighted only some of the non-overlapping criteria.

while those closer to the core, such as information exchange, are both harder to measure, and difficult to assess in a domain-independent and non-partisan way (one person's information may be another's rumor). Indeed, information and learning are often taken to be the paramount desiderata of deliberation ???, but this surely is not the core either, since otherwise we could sidestep the expense and merely sit citizens down with some pamphlets and a quiz with payments for correct answers.

But if there is some core process of deliberation that is of value because it leads to better democratic decisions, but which may be assessed independent of those outcomes, then it must be something more fundamental than amassing or expressing the right facts or emotions. Interlocutors must exchange facts and ideas, but with reflexivity and responsiveness (Habermas 1994, Wilhelm 2000), not just providing information, but providing information and ideas that are relevant to what their interlocutor has said. Neither repeating their own hobby-horses nor echoing the other, at the very least they should say something that is new and relevant – relevant in the sense that it is somehow conceptually connected to what has just been said. At its core, public deliberation should reflect its namesake, the deliberation of the individual: the collective not just gathering information, but considering, exchanging and recombining it in an effort to reach the best decision.

This process is some ways is much more stringent than the other criteria: people must speak in detailed response to what others have said, and not just promulgate their favorite facts. On the other hand, it is still just one aspect of the complexity of deliberation, and by itself would probably not entirely satisfy any deliberative theorist. But even so, we argue that it is closer to the core process of deliberation, and often overlooked or at best indirectly measured – in large part because it is so hard to model in a non-domain-specific manner. Nevertheless, we argue that this responsive deliberation can be modeled and measured, and far from being a delicate flower that must be coddled with elaborate and expensive polls, it is robustly prevalent even in the rough-and-tumble of online political argument.

To test this claim requires a multi-part empirical model of responsive deliberation: We

must have large numbers of recorded political exchanges over time; we must have a method for extracting the underlying ideas, topics and concerns from this text; we must infer how these ideas and concerns are psychologically interconnected in the minds of the discussants; we must have a testable criterion for deliberative responsiveness that may or may not be met; and we must show that all this talk appears to have some actual effect on speech and opinion. To do so using observational data requires content that is much richer and more like day-to-day conversation than the potshot and echoing worlds of Twitter or Facebook ???. but such forums are surprisingly abundant elsewhere, and we focus here on the largest two political forums on the left and right. And because both allow users to like or recommend each other's posts, we can also examine the complex interplay and tension between ideological behavior (liking), speech, and deliberation, to show how deliberativeness varies both within and between ideologies. While we cannot assess every aspect of deliberation here, we present a specific and powerful method for assessing an important aspect of deliberation, one that can be applied to domains online and off, and which sheds light on a number of theories from political psychology, framing, deliberation, and persuasion.

The structure of the paper is as follows: In section 2, we discuss the structure of the online forums we examine, and how the ability they grant users to like each other's posts reveals latent ideological dimensions, and allows us to examine both mirroring and continuity between the left and right. In section 3, we turn to the textual content of what they say, developing a topic model not just of what people write, but of the interconnections between topics, which we interpret as a psychological network of ideas. These topics and connections vary by individual, forum, and ideology, and reveal how ideas, emotions, and speech differ both across and within parties. In section 4, we apply this psychological network model to online dialogues, showing how our model of responsive deliberation – where interlocutors respond to earlier posts by offering up topics related to, but different from, what was said before – explains many of these exchanges better than simpler models of expressive speech or framing. Lastly, in section 5, we leverage the longitudinal structure of these communities to test (as best as is possible using pseudonymous

observational data) whether all these discussions are having an effect, and find that this appears to be so, but that these effects vary in complex ways by topic and ideology.

Overall, we develop and test a rich but focused model of deliberation in the wild: Even in these fiercely ideological online worlds, users reveal complex psychological networks and short-term interactions that are responsive to what their interlocutor has said, and which appear to have long-term effects not just on speech, but on ideology. Most intriguingly, these psychological structures, short-term interactions, and long-term effects – which together comprise a deliberative model – vary in important ways with the ideologies of the speakers and their forums, pointing the way to a more feasible form of deliberative engineering that leverages the deliberative tendencies already in place online and off.

## 2 Forums, "likes," and ideology

There are no large bipartisan online forums, apart from the comments sections of some newspapers and other mainstream news media. The reason for this can be seen in those comments sections: generally they are filled with vitriol and ad hominem partisan sniping. Few users stick around for more than the drive-by accusation, and fewer still are engaged in anything like a dialogue. "Never read the comments" has become something of a watchword on the internet. However, this dismal political picture is far less accurate for unipartisan online forums, which in addition to being active and influential, are – we will argue here – far from the echo chambers they are often portrayed to be.

Dailykos.com is the largest political forum on the internet, with hundreds of thousands of active users, many of whom are active donors and activists, and all on the left. Redstate.com is perhaps the largest political forum on the right, with fewer users, but still many thousands daily. Dailykos runs its forum on its own platform, but Redstate uses Disqus, general-purpose discussion platform that has over 500 million users across thousands of websites – a size comparable to Twitter or Facebook. Like many forums, both Dailykos and Redstate share a similar structure: front-page posts by featured writers, along with a separate list of "diaries" by individual members;

each main post then has a threaded discussion beneath, and users can moreover vote to like or recommend each other's posts.[2] These are very politically engaged individuals, even if they number in the hundreds of thousands – a meso level of high information potential activists, in between the traditional political elite and the low-information voters. Figure 1, for instance, shows the prevalence of a few key words over 2009-2010 on Dailykos: clearly visible are topical surges of interest in the public option, town hall discussions, the BP oil spill, and Afghanistan. For the following analysis, 125,000 posts from Dailykos dating between January 2009 and December 2010 were collected, and 75,000 posts collected from Redstate dating between January 2013 and April 2014.[3]

Because we are interested in specifically *political* deliberation, we need a measure of political ideology in addition to our examination of the purely communicative behavior. While this could be extracted from the text itself (Laver, Benoit & Garry 2003, Slapin & Proksch 2008, Beauchamp 2011), ideally we would like a measure of political behavior and ideology that is separate from what users say, so that the two may be more independently compared. But happily, because users can "like" each others posts, there is a ready-made vote-like matrix that can be extracted from these forums and used to scale users along potentially ideological dimensions, just as the vote matrix in Congress can be used.[4] This vote-like data was scraped from each site for the 1000 most active posts,[5] and users and posts were scaled in a joint, unidimensional space (separately for each forum). One intriguing question is simply: what do we expect the main dimensions of disagreement will be *within* each partisan forum?

Table 1 shows a selection of diary titles from either end of the inferred spectrum (first prin-

[2]Dailykos is unique in that these recommendations directly shape which diaries appear on the front page, whereas on Redstate, the likes only affect the order of sub-posts beneath the main post, and only then if the user has opted for that view.

[3]These dates were chosen mainly for convenience: Dailykos switched to a less scrapeable format shortly after this time period, and Redstate only shifted to Disqus in late 2012. However, both periods were interesting times for their respective parties.

[4]A variant of principal component analysis is used here, but Beauchamp (2011) shows that this produces nearly identical results to the popular NOMINATE vote-scaling method (Poole & Rosenthal 1985, Poole 2005). In this case, the vote-like matrix is superior in some ways to even the Congressional record, since unlike voting records, votes for posts are much less public and thus more likely to be expressions of genuine personal preference.

[5]For Dailykos, the 1000 most recommended diaries were used; since diaries are not like-able on Redstate's Disqus-based platform, the 1000 most liked posts within any thread were used.

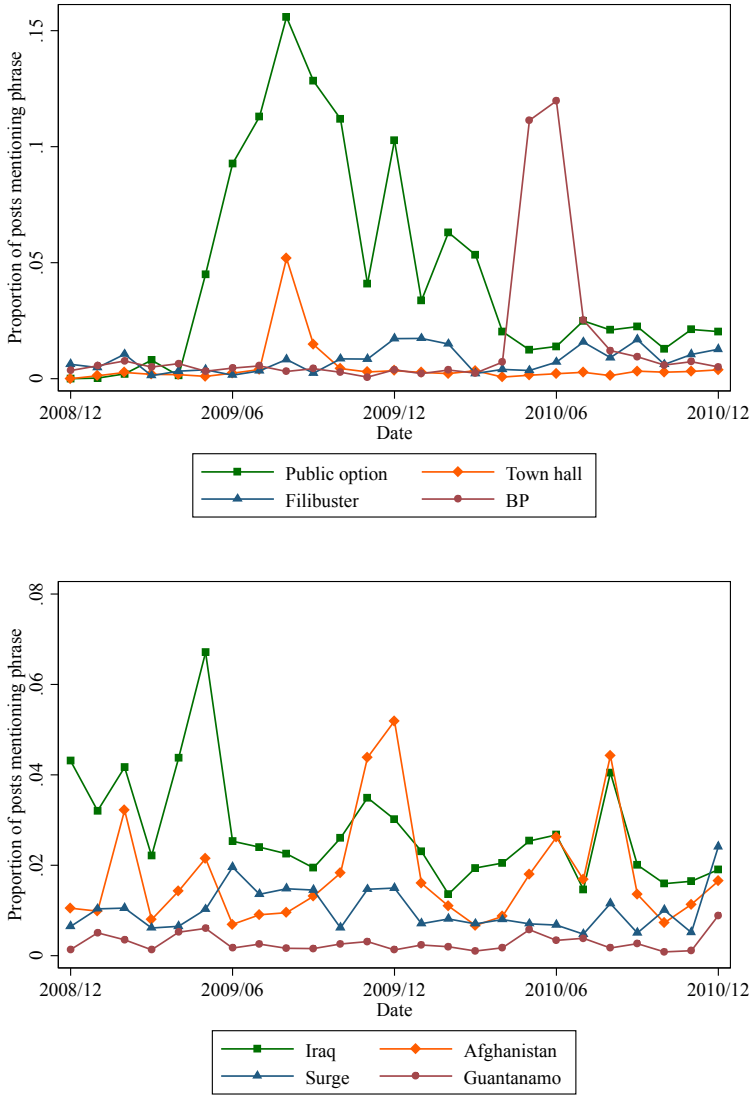


Figure 1: The frequency of topical phrases in Dailykos.com comments over the period 2009-2010. Top: Health and BP oil spill. Bottom: Foreign policy and war. Note the second bump in "public option" due to the Senate debate and filibuster, and the bumps in "Afghanistan" during the surge and Obama's August 2010 pullout announcements.

cipal component) for Dailykos; since the Redstate posts lack titles, Table 2 shows the first words of a some selected posts from either end of their dominant spectrum. For Dailykos, the primary dimension of intra-party disagreement is surprisingly recognizable, revolving clearly around praise or criticism of Obama.[6] The criticism is largely from the left – and indeed, many of the diaries

[6]Note that those diaries that score in the middle are very different, and much less recognizably ideological:

are policy issues not directly connected to Obama – so we have labeled these sides as Left and Center-left, but of course that is just an interpretation, especially since the clear unidimensionality of the two-party US system breaks down within parties. Arguably, the "center-left" here is less about ideology than about personality and emotion, full of praise lavished on the newly elected president.[7]

Given this structure within the left, one of the recurring questions here is what we expect the structure on the right to be: the same but shifted, or mirrored (with the personal towards the center again), or something else entirely? Table 2 shows the leading lines of the highest-scoring posts on Redstate, and one group is immediately recognizable: Rand Paul. The other is more like the left wing on Dailykos, full of explicitly ideological language castigating the left, extolling guns, and decrying the "inner cities" and blacks. But the overall structure is similar, almost mirroring Dailykos: personal and laudatory on one side (which we have labeled centrist, though that is of course arguable), and more stridently ideological at the extreme.

Thus even within these unipartisan forums, there are clearly dimensions of disagreement: the vote scaling shows users tend to vote for one side or the other, but rarely both. This would seem to bode poorly for dialogue and deliberation, replicating partisan divides on a fractal scale within the parties. But as we will see, when we turn to the text of what people write, the picture is not so dim – there are strong preferences, but they don't preclude strong interactions.

But before turning to the content analysis proper, the vote-scaling approach does allow us to say something about how people write. That is, do they tend to write posts that fall on the same ideological side as the posts they like? The scaling method scales both users and posts (with no knowledge by the algorithm of who wrote what), so correlating writer ideology with post ideology directly addresses this question. On both forums, the overall correlation between the author's position and her texts is strong, at around 0.4. But this too varies by ideology: On Dailykos, the

On Dailykos, they deal with non-divisive topics such as helping Haiti victims, sending CDs to Iraq, odd science discoveries, the perfidy of Republicans, etc.; on Redstate, they deal with more political topics, but again ones without disagreement: the sins of Obama, liberals, the media, etc.

[7] Although this attachment to the person of the president may also reflect theories of the authoritarian right (Adorno et al. 1950, Altemeyer 1998).

Table 1: Selected center-left- and left-leaning diary titles from Dailykos.

| |
|---|
| *Center-left* |
| Boycotting *This* president? What a disgrace. What a shame. |
| OBAMA EATS REPUBLICANS' LUNCH! |
| Already one of the GREATEST PRESIDENTS EVER. |
| The Obamas: Making the world a better place, one day at a time |
| *Left* |
| Barack Obama is out of touch. |
| The White House Lashes Out At Howard Dean |
| This Is Why I Don't Donate To The DNC! |
| No Public Option? I'll See You In Court. |
| What Happens if Obama Loses the Left? |

Table 2: Selected center-right- and right-leaning diary titles from Redstate.

| |
|---|
| *Center-right* |
| Yesterday was an important day for our party in that Paul is showing the way to ... |
| I liked the part where Sen Paul said to the Senate ... |
| The smartness of Rand Paul... |
| I will never understand why Arizona re-elected McCain this last time around.... |
| I have whole new admiration for Rand Paul. Finally a Republican who ... |
| *Right* |
| Marxists, Socialists, Democrats, Lefties, Left of Center, Far Left... |
| Good luck with that. Out here we don't have marxist policies against gun ownership.... |
| When the inner cities melt down, those of us in Flyover Country won't shed a tear.... |
| Conformicrats, statist, authoritarian control freaks.... |
| A universal background check on the GOOD guys will do nothing except give the govt a LIST... |
| No. Young black males are a danger to public safety. Deal with the problem, not the excuses. |

centrists have much higher correlations between what they write and what they like ($\rho = 0.40$) than the far left ($\rho = 0.11$); conversely, on Redstate, the far right is much consistent between what the write and what they like ($\rho = 0.71$) than the Rand Paul centrists ($\rho = 0.07$). This breaks the earlier mirrored symmetry, suggesting that while both groups have centrists who are more personality oriented and less explicitly ideological, there appears to be a shared tendency for the relative right to be more doctrinaire in what they like and write than the relative left.

Given these similarities and differences, how would we expect these ideological factors to affect how these users write and respond when they do choose to interact with others on their

forum? To answer that requires going beyond these simple vote-based structures to examine the text itself. One major limitation of many existing studies of communication on social networks is that they limit themselves to likes, retweets, and other simple structural measures. While these are often ideologically-shaped behaviors, they are distinctly limited in what light they can shed on the content of communication. The forums analyzed here differ from Twitter or Facebook in the richness and length of the communications (amounting to many gigabytes of text in our corpus), but modeling this text requires a more complex toolkit.

## 3 The topic model as a structure of ideas

Although it is possible to scale speech in much the same way that votes are scaled, using a matrix of users and words frequencies instead of users and liked posts, this is really only useful when the results are dominated by a few latent dimensions. For instance, in domains such as US legislatures, speech is dominated by partisanship, and speech scaling reveals clear oppositions between Democratic and Republican language (Monroe & Maeda 2004, Slapin & Proksch 2008, Beauchamp 2011). But although the voting behavior on Dailykos and Redstate may be dominated by a few latent dimensions of disagreement, such is not the case for speech, which instead shows clusters of topically related words rather than clear ideological dimensions.

This tension between low-dimensional ideological tendencies and high-dimensional speech is at the heart of the deliberative challenge, resisting the impulse to settle into a repeated subset of talking points and favorite arguments. As such, it is essential to capture the more multifarious nature of speech, and we do so here using a modified version of the correlated topic model, which in turn is similar to the popular Latent Dirichlet Allocation (LDA) topic model (Blei, Ng & Jordan 2003, Blei & Lafferty 2007). The LDA model itself has now been applied to a variety of political domains,[8] so the core idea will only be briefly discussed here as it pertains to the

[8] Although in many of these cases, LDA is applied without much reflection about why an unsupervised topic model is preferable to a supervised model that is trained on the categories the researcher already knows and expects. In the present case, by contrast, the main substantive interest in these intra-party disagreements is precisely that we do not, a priori, know what it is these political differences will look like.

present purposes. LDA belongs to a general class of generative Bayesian models, which in this case imagines that documents are generated stochastically and without regard to word order: each document has a distribution over K topics, and each topic has a characteristic word distribution. Thus a document is taken to be generated by drawing from the topic distribution, and then each word is taken as a draw first of a topic from the document's topic distribution, and then of a word from that topic's word distribution. (See the Appendix for more technical details about the modifications of this model for the purposes here.) Crucially, as with the correlated topic model of Blei & Lafferty (2007), we can also examine the topics for correlations within the corpus; this network of correlations will form the essential conceptual structure used to predict deliberative exchanges between interlocutors.

Table 3 shows the results of estimating a 10-topic model[9] of the 125,000 comments analyzed from Dailykos; for each topic, the 10-15 highest loading words are shown, to illustrate the content of those topics. Many of the topics during this time period (2009-2010) are immediately recognizable: health (1), class issues (10), party politics (9), Obama and the public option (6), war and foreign policy (8); a couple of more emotional topics, one positive, one negative (3, 5); and two more topics that are more inward looking and meta, discussing the diaries and posts themselves (4, 7). The numbers assigned are arbitrary, but the order in Table 3 is not: the loading for each topic in each post can be correlated with the ideology of the post as measured by the earlier scaling, and Table 3 shows the topics ranked from most left-associated (top) to most center-associated (bottom). What is striking about this ordering is how it reveals two sides that differ not just by ideological prediliction, but also by content: the policy topics are dominant on the left, while the emotional and meta topics are dominant at the center (relatively, the right). This is consistent with the diary titles shown in Table 1, but presents the psychological structure in much more detail: policy and political terms on one side, emotional and meta-argument on the center / relative right.

[9] 10 was chosen both for legibility, and tractability in the later sections, where parameter estimates grow as the square of the number of topics.

Table 3: Top words for the 10-topic model of Dailykos, ordered from far left to center-left.

| | |
|---|---|
| 1: | bill, insurance, care, system, health, financial, america, companies, economic, money |
| 10: | people, jobs, class, money, tax, war, country, world, military, black, racism, middle, power |
| 9: | house, party, democrats, money, white, democratic, vote, line, senate, republicans, time |
| 6: | public, people, obama, care, option, health, change, trying, reform, saying, doing, single |
| 2: | government, president, americans, oil, private, look, based, world, day, million, post |
| 8: | obama, president, bush, torture, war, administration, rights, congress, iraq, vote, law |
| 3: | wrong, hate, getting, please, win, hope, guy, people, little, mind, fucking, bad, day, doing |
| 5: | people, actually, life, love, family, own, lot, help, little, feel, social, time, believe, god, sorry |
| 4: | time, people, left, news, diary, story, called, read, personal, agree, days, political, sense |
| 7: | diary, person, diaries, comments, people, comment, support, thanks, diarist, fdr, kos, paid |

*Note:* These are the highest-scoring words from $\beta_{k,n}$.

Having combined the topics with the ideological measure, the final layer is the connections between topics. Figure 2 shows the correlations among these topics – which topics tend to co-occur in posts.[10] Red links are negative correlations, green are positive; the color of the nodes correspond to their ideology (blue = left; red = center/right); node and link sizes are by eigenvector centrality. This paints the fullest picture not just of the structure of debate on the forum, but, we argue, the conceptual structure in the minds of the users relating these various topics and concerns. Thus 6 and 9 are tightly connected, revealing the twin concerns about the health bill and the party negotiations dominating this period; and 1 and 10 are also positively correlated, revealing the connections between class issues and the monetary pragmatics of the health bill. The emotional topics 3 (negative) and 5 (positive) are also intriguingly connected, perhaps because one political side (the centrists) tend to use both more, and these topics are anti-correlated with the more policy-oriented issues on the far left. Thus in one simple network a complex conceptual structure is revealed – albeit one with many possible interpretations, which themselves require deep contextual knowledge.

The picture painted on the right, on Redstate, is not quite so clear, and interestingly differs

[10]These are specifically the correlations between $\eta_d$ distributions, ie, prior to transformation to proportions. This ensures that the correlations are not all negative due the normalization to the simplex; see the Appendix for details.

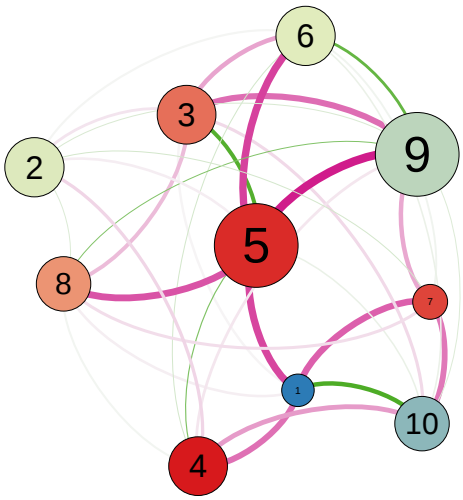


Figure 2: Dailykos topic network. Green links = positive correlation; red links = negative correlation. Red nodes are center / pro-Obama, while blue nodes are left / critical of Obama. Node size reflects eigenvector centrality, while link size reflects the correlation absoluate value. Force atlas network layout genereated with *Gephi*.

from the left. Emotion and meta words (post, comment, diary) appeared mixed throughout the topics, and in general the stucture appears more casual and less policy oriented. This in part reflects the more chatty environment of the Disqus commenting system, and in part reflects the sparser data from Redstate – but clearly also reflects a different sociological and conceptual structure on the right, mixing policy and emotion (mainly invective) more uniformly. To a lesser degree, as with Dailykos, we also see more partisan political topics on the extreme, and more personal discussions towards the center: note that four of the five bottom topics each have a specific name associated with them, for instance (Paul, McCain, Boehner, and Cruz). But overall, there is again less mirroring, and a bit more resemblance between Redstate overall and the centrists on Dailykos.

Finally, examining the correlations (not shown), the most positively correlated are 1, 2, 6, and 7 – again reflecting the tighter coherence among the more rightward subcommunities in each forum. The strongest negative correlations are between 5 and 8 vs 6 and 7 – the libertarians vs the rightwing partisans.

Table 4: Top words for the 10-topic model of Redstate, ordered from far right to center-right.

| | |
|---|---|
| 9: | post, site, agree, yes, law, maybe, money, asshat, reason, matter, establishment, candidate |
| 7: | party, bill, republican, country, stop, spot, support, hear, government, conservatives |
| 1: | people, issue, love, democrats, mind, thank, vote, commenting, apparently, talk, political |
| 2: | god, lost, tell, aren, doing, idea, world, wait, fight, cowards, real, write, tax, disagree |
| 3: | obama, care, left, yep, door, win, cowardice, choice, amen, thread, run, guys, liberal |
| 10: | comment, thanks, idiot, disqus, paul, own, called, troll, ok, bad, funny, mitch, libertarians |
| 6: | oh, redstate, diary, look, welcome, john, house, try, little, pretty, majority, mccain |
| 4: | yeah, boehner, read, gop, else, comments, guy, crap, idiots, ohio, word, stupid, church |
| 5: | exactly, won, leftist, vote, call, correct, banned, history, believe, true, stupidity, libertarian |
| 8: | bye, time, gone, actually, wrong, sorry, question, conservative republicans, cruz |

*Note:* These are the highest-scoring words from $\beta_{k,n}$.

To illuminate one last difference between the two sides, although we use the collective network in the following sections, we can also calculate topic networks for individuals who have posted sufficiently. These too can be examined for structural differences between and within the two forums, potentially shedding light on how these network structures differ with ideology. To test this, individual topic networks were constructed for the few hundred users in each forum with more than 50 posts; each user's network was then assessed for a variety of graph-theoretic qualities,[11] which can then be correlated with ideology. What we find is that there appears an intriguing connection between ideology and individual structure on the left, where a half-dozen measures (similarity, connectivity, edge connectivity, degree, edge weight, number of edges, density) correlate much more strongly with users on the far left, while measures such as path length and radius correlate more with the center (relative right). Figure 3 shows two examples that illustrate the general structural characteristics underlying these various measures: the networks on the far left are in general more densely interconnected, while those from the relative right

[11]Specifically, the 23 measures were assessed: Similarity, connectivity, edge connectivity, degree, edge weight, number of edges, density, transitivity, clique modularity (binary), number of nodes, knn, strength, eigenvector centrality, abs. weight, closeness centrality, modularity (max), diameter, betweenness centrality, degree centrality, path length, radius, independence. Weighted edges were converted to binary edges by using a Holm-corrected p-value threshold for the topic correlation significance. For node- or edge-level measures, the graph mean was used. The list above is sorted in order of correlation with left-ness on Dailykos.

(center-left) are both sparser and more spread out.[12] Yet once again, turning to Redstate, these differences are much less pronounced; while it is difficult to directly compare measures across forums (because of the different quantities of data), the left-right difference within Redstate is largely non-existent for these measures, with a small ($p < 0.10$) tendency for the center-right individual networks to have higher mean centrality scores, which also reflect more interconnected networks. So once again, the pattern – here, of diminishing interconnection as you move rightward – is continued, not reflected, moving from the Democratic forum to the Republican one.

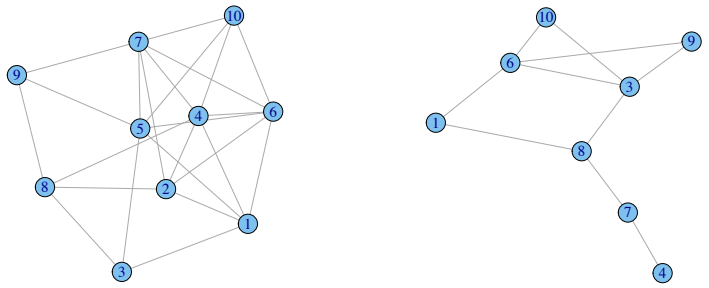


Figure 3: Examples of individual topic networks from users on the far left and on the center-left for Dailykos.

To summarize, while we do see more personal topics towards the center for both forums, the Democratic forum also shows a segregation between policy speech on the far left versus more emotional and meta speech towards the center, while the Republican forum has this emotional and meta speech spread much more uniformly, with only a light tendency for more partisan politics at the far right extreme. Rather than a mirrored relation between centrists and extremists, the pattern within the left extends straight into the right, with increased emotion and meta talk on the right of the left-wing forum, and then across almost the whole of the right-wing forum. And similarly, the pattern of diminished interconnections within individual networks continues from left to right across both forums, although we cannot directly compare relative levels between the two forums.

At this point, we have shown a rich communicative and conceptual structure underlying both

[12]These results remain even when controlling for post frequency, which is somewhat higher on the far left, and indirectly affects connectivity by providing more data for these users.

these forums, but the correlations and anti-correlations between topics and ideologies also suggests that these could be communities in name only, riven by the same sorts of disagreements and differences separating right from left. There is certainly plenty of opportunity for rich deliberation – but does anyone take it? To answer this requires not just a static model of speech, but a dynamic model of how people respond to each other in the give and take of post and reply.

## 4 Deliberative arguments

The core hypothesis here is that the network of topics as in Figure 2 represents not just statistical correlations among issues, but a conceptual structure that in a sense is shared among all participants: Some may prefer the blue nodes, some the red, but all agree about which are positively or negatively correlated. This conceptual network reflects the structure of thought itself: one hears or considers an idea, which in turn reminds one of associated ideas, and so on through the associational network. Thus a deliberative conversation is, as the word suggests, one that most resembles deliberation – ie, individual thought. An honest, responsive discussion consists of interlocutors who do not merely repeat their own concerns or slavishly echo each other, but rather, those who offer ideas (whether or good or bad) that are associated with, but different from, what has come before.

To test this model of deliberation as interactive deliberative thought, we examine pairs of comments, where comment B is a response to comment A. What do we expect the topics of comment B will be, given that it is responding to A? At the most skeptical, we might hypothesize that a response is utterly unrelated to the previous comment or the ideology of the speaker. More reasonably, according to the behavioral models discussed in the introduction, we might expect that, whether they are engaged in framing, agenda-setting, priming, or even simple expressive position-taking, users will express their favorite topics and concerns regardless of audience, beating their personal drum whatever the recipient. A third, more responsive model might expect that naive subjects would simply adopt the topics of the post they are responding to – a position perhaps slightly less naive given that users choose who to respond to.

By contrast, the deliberative network-of-ideas model suggests that respondents will deploy ideas that are related to, but not the same as, the ideas promulgated by the initial writer. To quantify this hypothesis, we can interpret the topic network as a markov matrix, where each entry $P_{i,j}$ (the links between topics) is the response of topic $i$ to topic $j$: the reader encounters a comment with topic distribution $A$, and these topics prompt a response with distribution $B$, where the topics $B$ are those that were connected to the topics in comment $A$, such that $B = \mathbf{P}A$. To estimate $\mathbf{P}$, we simply regress each topic $i$ on $j$ – ie, this is essentially equivalent to the covariance matrix that generated our network of correlated topics, but suited to the markov process.[13] Of course, in actual practice, a post is not raw topics, nor is the response: in reality, issues under the umbrella of topics are raised, and other specific ideas, facts, and points are then raised in response; the topic response model here is once again of necessity a simplification that hopes to capture the coarse dynamics of a deliberative interaction.

As an illustrative example, see Table 5 and Figure 4, which show a comment and response from Dailykos.[14] In Figure 4, the first comment has topic distribution shown in green, while the response is in blue. The network model prediction $\mathbf{P}A$ is in red. A left-wing critic of Obama A primarily raises topics 1 and 10 (which are associated with Obama criticisms over the health bill and the economy) along with some 3 (strong language), and the response B sticks with 1 (health) but switches tone (no 3) and moves mainly to 9 (legislative and party politics). The network-based prediction generally anticipates the direction of these switches, but although the results are substantively interpretable, because the network utilizes the entire set of topic interconnections to make the prediction for a response, one cannot simply visually inspect Figure 2 to guess exactly which output topics will be predicted.

The fundamental advantage of this algorithmic approach, however, is that it can be scaled up far beyond what is feasible with hand-coded transcripts. We can now test our deliberative

[13] More specifically, the entries of $P_{i,j}$ are the coefficients from bivariate regressions, not a multiple regression of topic $i$ on all other topics $\mathbf{j}$; the constant terms are dropped (most are insignificant); and the matrix is not normalized.

[14] Rather that proportions, the topic values in the log-transformed space are used, since this is the space in which the covariances and regression coefficients have been calculated. For the example in Figure 4, the predicted values were then transformed back to proportions using the inverse centered log transform.

Table 5: Excerpts from a Dailykos comment-response pair.

| |
|---|
| **A**: I'm fine with drawing contrasts against Repubs in terms of what we've done so far. But a large portion of what Dems plan to do in November is to run against the other guy, to say we can't let Republicans win in November. What we've done so far—in terms of stimulus, health reform, etc, really haven't gone far enough... |
| **B**: Oh, we just chose not to make the stimulus bigger. It had nothing to do with the fact that there weren't the votes for a bigger stimulus, which was because of conservatives. So would you have preferred no stimulus at all? But like I said...we can't run on anything because we haven't done anything. And you don't want them to point out what a Republican takeover would mean. So, what the hell should we run on?... |

*Note:* Some spelling errors and typos corrected.

model against the various null hypotheses discussed earlier: Do interlocutors respond randomly, or merely echo the first speaker, or simply beat their own drum? Or do at least some of them engage in these more responsive, deliberative interactions that raise points related to, but different from, what has been said before, in the way we expect argument to ideally work?

The results of this test are shown in Figure 5, and constitute the core finding of this paper.[15] For each comment-reply pair, we predict the topics of the reply based on the specified model, and take the total absolute error between that prediction and the true distribution for that reply. For each method, the mean absolute error (MAE) is calculated, and then because these are pairwise tests between methods using the same data, each line shows the difference in MAE for the specified test versus the baseline, which is simply taking the mean topic distribution for the forum as our guess for any reply. The error bars are 95% confidence intervals on this difference; larger numbers are better, and being to the right of dotted line means a model has outperformed the baseline.

As can be seen, for both forums, the random-reply model performs the worst, and the hypothesis that replies simply echo the previous comment, while significantly more accurate than random, also does quite badly. After that, the two forums somewhat diverge. On Dailykos, of the relatively effective models, the next best is assuming that the user always replies with her usual set of topics – but guessing the forum mean does even better than that, suggesting a fairly high

[15]Results are shown for the full sample, but out-of-sample cross-validation produces similar results, because the matrix $\mathbf{P}$ is not very sensitive to subsample variation.

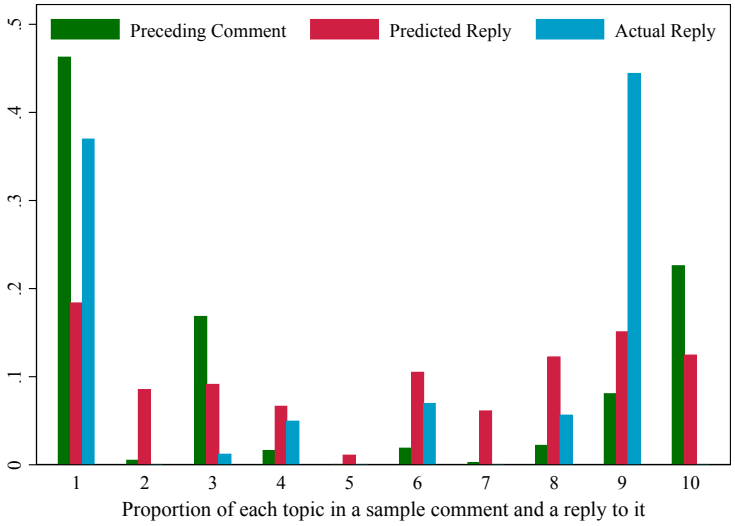


Figure 4: Topic distributions for a comment, a reply, and the network-based prediction. (Comments shown in Table 5.) The predicted topic distribution of the reply is reasonably close to the actual distribution, shifting from an anti-Obama attack to a pro-Obama defense.

level of forum coherence. However, more accurate than either is the network prediction, and this is further (and significantly) improved by combing the network prediction with a mixture of the writer's mean topics plus some of the previous comment's topics.[16] Thus on Dailykos, the best prediction for a comment reply is the network model of deliberative response, where interlocutors offer these topics that are related to what was said before. This is a strong validation of both this network/markov model of discussion, and the deeper model of interactive deliberation.

Turning to Redstate once again reveals a related but importantly different pattern. Here, guessing the author's mean does a better job than the forum mean, and both do better than the network prediction, suggesting a community that is less cohesively similar, and given to greater degrees of non-responsive user repetition of favorite talking points. However, a mixture of the network and author mean surpasses the author mean by itself, suggesting that there is a deliberative element, but that this responsiveness is much more diluted by author-specific drum

[16]These results are not sensitive to the level of mixture, and the previous comment contribution in particular can be omitted with little loss.

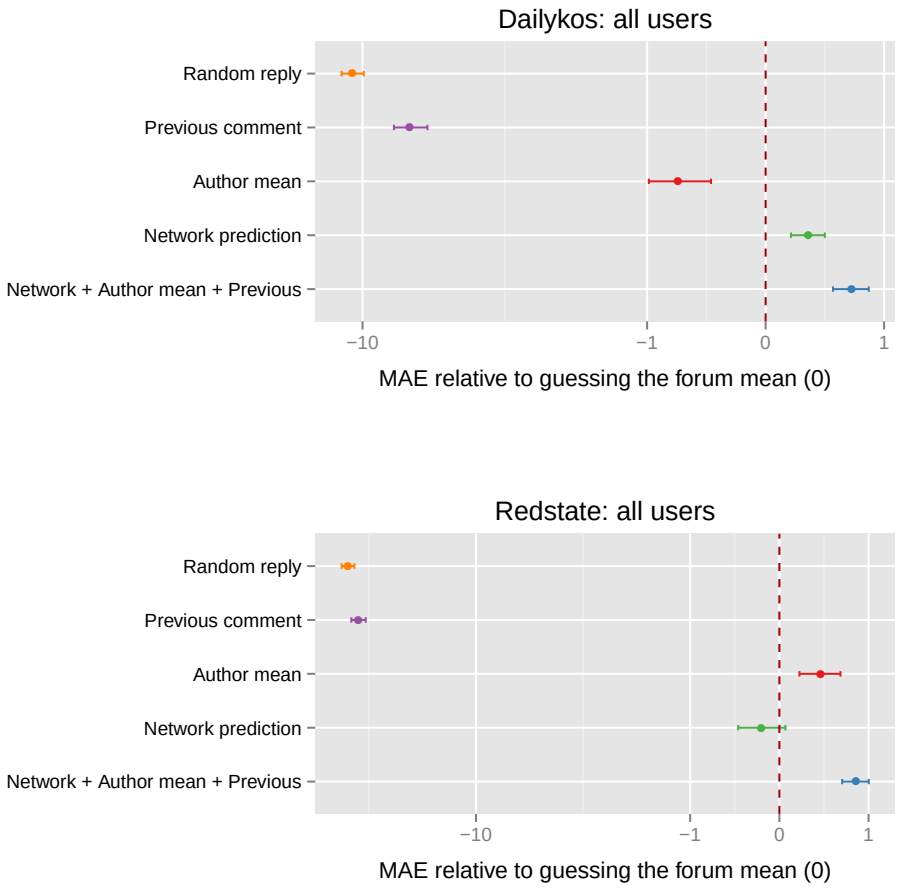


Figure 5: The accuracy of the various models predicting the topics of response $B$ as a function (or not) of initial comment $A$.

beating.[17]

In sum, we see a strong validation for the deliberative model on the left, and a somewhat weaker validation on the right. This validation supports our model of networked concepts, and shows that a responsive dialogue does exist in the wild, where interlocutors raise ideas related to what the other is saying. Far from being a domain where everyone screams their own concerns or merely cheers each other along, we instead see that at least some of these users must be engaged in responsive interactions that resemble a fundamental aspect of deliberation. While this does not demonstrate that all of the criteria for deliberation are met, it does show that one of the

[17]These models models can also be run separately for the right and left ideological halves of each forum, generating for each its own network $\mathbf{P}$. But the results of these separate tests are largely the same as for the full forums, suggesting that the relative left and right networks within each forum are not deeply different.

most important ones seems to be met, at least for some of the users – without a purposefully crafted deliberative environment, and in some of the least promising (by reputation) corners of the internet.

This leaves one final question. Do these short-term interactions have any significant long-term effect, or are they just brief, evanescent encounters?

## 5 Long-term opinion change

We have described a new model of deliberation and political psychology that can be inferred wholesale using the topic modeling approach, and the short-term predictions of this model were tested and shown to improve upon behavioral theories that imply either that discussants will only emphasize their own frame or only echo the frames established by others. But as Lasswell famously put it, the basic questions in studying interpersonal communication are "who says what to whom with what effects?" (Lasswell 1948). Does any of this have any effect? Deliberation without effect may look nice, but clearly fails at one of the fundamental requirements.

Because this is observational and not experiment data, the causal question of persuasive effects can never be answered definitively, but once again we will do the best we can given the limitations. The key advantage is that these are longitudinal as well as cross-sectional data: we can track individual users over time, and see how their behavioral changes follow on changes to what they read. Granger causality may be a misnomer, but time-series cross-sectional data at least allows us to control for many possible confounders on the individual level.

Alas, the Redstate data contain too few users posting abundantly over sufficiently long periods, so we must restrict our long-term analysis to the generally more robust Dailykos data, stretching over 2009 and 2010. Does hearing more left or centrist speech simply pull the listener in that direction? Does hearing more of topic X induce more speech subsequently containing topic X? How do the ideological aspects of the topics influence their persuasive effects? These issues are crucial for the deliberative question, since properly deliberative discussion should no more draw a group into a homogenous mean than it should leave them unchanged. Rather, as

with the short-term interactions, we would hope for something equally richly responsive to the complex topical and ideological structure here.

The basic issue, then, is to examine for each individual over time the topics they "hear" as input, and the topics they "speak" as output, along with the vote-based ideological measure (which also can change over time). Does variation in what is heard predict subsequent variation in what is spoken and how users vote? Individual voting behavior and ideology does indeed change over this time period for many users – 2009-2010 was a contentious time for Democrats, a time of much disillusionment. Figure 6 shows, for the 1000 most active Dailykos users, the coefficients of a regression of their ideological position calculated monthly against a simple time trend; we see that about 5% who begin on the far left trend significantly to the (relative) right, while an approximately equal number on the other side trend leftward. (These are many more trenders than one would expect by chance even after correcting for multiple testing.) There do indeed seem to be many users who shift from Obama defenders to left-wing detractors, and vice versa. Presuming that such changes have causes that are mainly external to the site even for the most dedicated users, can we find evidence suggesting potential internal causes as well?

To address this question, we take two different approaches, both of which seek to leverage the longitudinal structure to control for the individual fixed effect that we cannot measure. The first approach is a time-series cross-sectional approach using individual fixed effect, which with many interacting factors can be estimated either using an ensemble of bivariate models, or jointly using panel vector autoregression methods. The second approach follows the matching literature, where we match those who appear to encounter unusually high degrees of topical speech in period 1 with matched controls who appear very similar, and then look in period two for how the behaviors of those two groups diverge. Both approaches present complex pictures of how these myriad variables interact, but the fundamental question will be whether there are any apparent persuasive effects at all, and whether these vary with ideology and topics – which we will see they do, albeit in complex ways.

The first step is to decide what it is that users are perceiving within the forum. Since users

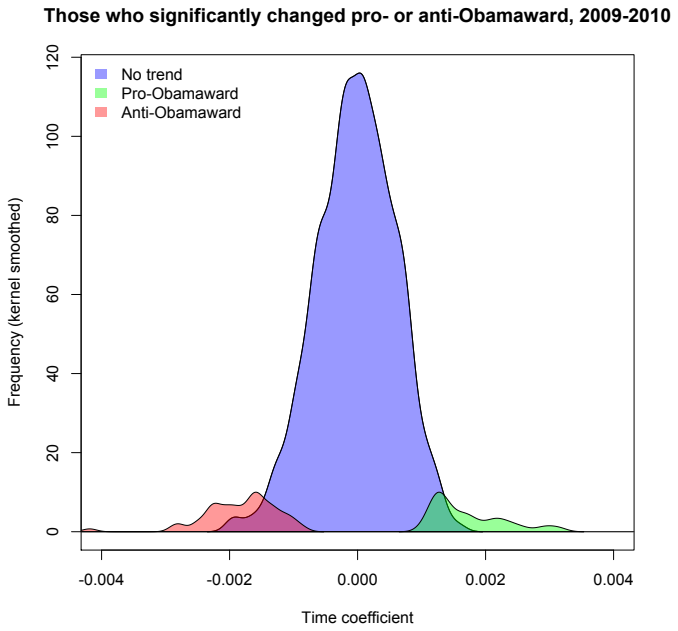


Figure 6: Dailkos users who trended significantly from right to left or vice versa over 2009-2010.

choose who to respond to, using the comments upstream from them that they are replying to runs perilously afoul of endogeneity, since those posts have been selected by the user in choosing to reply to them. A better, less endogenous input is the replies users receive to their own comments; although the previous section has shown that these too are affected by what the user has written, this effect is much weaker than what the user chooses to respond to, and thus is considerably more independent of the variables we take as the dependent variable, the user's preferred topics and ideology. There is also considerable evidence from A-B-A triplets that these users do tend to read the comments to their own comments – and at worst, if users never read their replies, that should only bias things towards a null result.

For the panel approach, we have 24 time units, aggregating the data by month, and 1000 users; we have 11 dependent variables (the mean topic expression for that user in that month, plus the vote-based ideological measure); and we have 10 independent variables (the mean amount of each topic heard in that month). Given that there are many interacting variables, the first

approach is simply to estimate the effect of each of the independent variables on each dependent variable separately, controlling for individual fixed effects to capture the relation between their variations. Thus the model is

$$y_{it,l} = \beta_m x_{it,m} + f_i + y_{it-1,l} + \epsilon_{it}, \tag{1}$$

where $l \in \{1...11\}$ and $m \in \{1...10\}$, and each bivariate model uses the Arellano-Bond (AR 1) estimator employing the lagged dependent variable as instruments to control for endogeneity (Arellano & Bond 1991).

The second panel approach is to use panel vector autoregression (PVAR), which explicitly allows all these variables to be interacting with each other, instead of the unidirectional causation implied in the previous model (Holtz-Eakin, Newey & Rosen 1988, Love & Zicchino 2006). Given that what people say does influence what they hear, this is a reasonable assumption; the drawback of the PVAR model is that it is dangerously subject to over-fitting, since one is essentially estimating a 21x21 variable model, and indeed twice this with the lags. For this reason, PVAR models are rarely run with more than a few variables, and indeed the existing package had to be revised to handle this many variables.[18]

Our PVAR model is thus:

$$\mathbf{y}_{it} = \sum_{l=1}^{m} \Gamma_l \mathbf{y}_{i,t-l} + f_i + d_t + u_{it} \tag{2}$$

That is, $\mathbf{y}$ is a vector of $k$ variables which are all functions of their own lagged values plus the lagged values of all the other variables, as well as unit- and time-specific effects.[19] One final complication

[18]The Stata package developed by Love & Zicchino (2006) was originally designed for up to six variables. Adapting this software for a much larger set of variables requires a number of non-trivial changes to be made, particularly in calculating the covariance matrices used to generate the impulse responses. The updated Stata algorithm plugin is available upon request, although it now also requires an R script (and associated packages) to generate the duplication matrices (Hamilton 1994) used in the Monte Carlo stage.

[19]To estimate this model, the data are first time-demeaned to eliminate time-specific effects. For the usual dynamic panel approach, fixed effects could also be eliminated by taking first differences before using the lagged regressors to estimate the model using GMM or 2SLS (Arellano & Bond 1991), but since the fixed effects are correlated with the regressors due to the lags of the dependent variables, first differencing would create biased coefficients. Instead, following Arellano & Bover (1995), variables are forward mean-differenced (i.e. the mean of all the future observations available for each unit is subtracted), and the untransformed variables are used as the

for the PVAR model is that one cannot simply read off the coefficients to get the input effects; one must instead consider the fact that a shock to variable X affects all the variables, and simulate impulse-response functions to get the overall effect of a shock to variable X on variable Y. But the final upshot is similar to the bivariate approach: for each of the 11 dependent variables, we have a (purported) effect of hearing more of topic X on that speech or voting behavior Y.[20]

This is obviously a huge variety of purported effects, but they can be summarized by the heat maps in Figure 7. The results of the two models appear different, but the heatmaps correlate at $p < 0.01$. Most importantly, in either approach, there are many coefficients that appear statistically significant, even after Holm multiple-testing correction. This clearly suggests that, *pace* the usual caveats about causal inference using causal data, there do appear to be close connections between hearing more of topic X in period $t$, and speaking more or less of other topics in periods $> t$. And overarching these myriad interactions – where each cell tells a potential story about how hearing more of topic X might cause more or less subsequent speech on topic Y – is a more general pattern, where positive (green) correlations appear more on-axis, and red more off-axis. That is, hearing more of left-wing topics tends to produce more left-wing speech in the future, and vice versa for more centrist (personal, emotional) topics. And looking to the ideology (the PCA of the vote), here too heard speech seems to affect one's deep ideological preferences, with left-leaning topics producing more left-leaning voting, and vice versa – although the two models disagree on which topics are most effective, perhaps reflecting their differing emphases on direct versus total effects.

The matching approach is considerably simpler. In the first period (the first year), we simply construct a "treatment" group for each topic, where each treatment group is the 10% of users who heard an unusually large amount of topic X, but are otherwise undistinguished. To

instruments. See Love & Zicchino (2006) and Holtz-Eakin, Newey & Rosen (1988) for further details.

[20]The responses of received topics to spoken topics are also estimated in the PVAR model, but these are less interesting for two reasons. First, because as the previous section showed, responses received are a direct function of what was said. And second, for technical reasons the panel vector autoregression model assumes that variables do interact contemporaneously, but only in a set order, where an earlier variable in that order can affect a later variable but not vice versa. This requirement is well suited to the actual state of affairs here, where we would expect that ideology affects speech which in turn affects received speech in a given time period, but that causation only flows the opposite way over longer periods of time. The result is that shocks to speech naturally, but uninterestingly, show strong effects on received speech, just as we would expect.

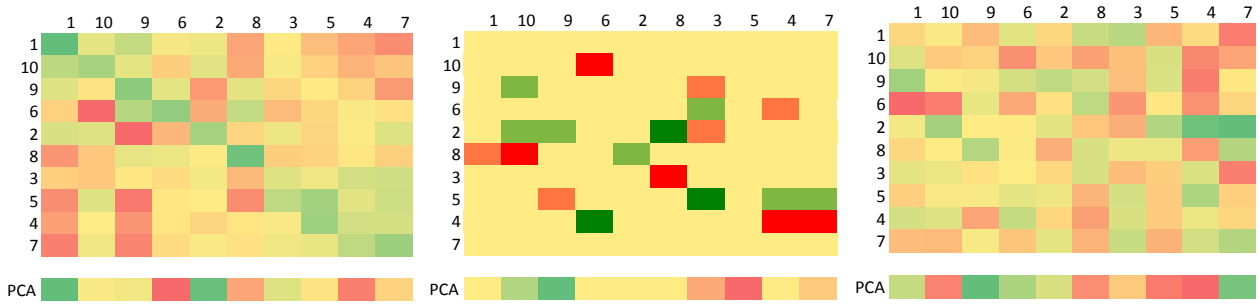


Figure 7: Heatmaps of the effect of hearing more of topic X (top) on speaking more of topic Y or voting (left). Green=positive, red=negative, yellow=not significant. Right: bivariate models; center: PVAR model (effects from impulse-response simulations); right: difference-in-differences from matching. Topics ordered from left (top) to right (bottom).

this treatment group are matched the same number of pseudo "control" users whose distribution across the other 20 variables is identical (the 9 other heard topics, plus the 10 topics spoken and the voting variable) – these are users who appear very similar to the treatment group, apart from happening to hear less of topic X. We then compare each treatment-control group pairing in period 2 (the second year) by taking difference-in-differences, and see which of the 11 outcome variables now diverge between these two groups, which ideally would control for all individual-level confounders between these two groups. The result can again be expressed as a heat map, where green cells are difference-in-difference measures where the treatment group shows significantly more of output Y than the matched control group in period 2. The results are shown in Figure 7. Once again, though it is hard to visually discern, this heatmap does significantly correlate with the other two ($p < 0.01$); far more of the cells are significant than would be expected by chance even after multiple-testing correction; and a $\chi^2$ confirms that there is a general tendency for more left-leaning topics to produce more left-leaning speech, and vice versa.

Again, it is impossible to prove that these effects are causal, and there are many reasons to worry that any systematic tendency to hear more of topic X following by saying more (or less) of topic Y may be due to external sources that affect both speaker and listener. But the general picture does appear to be one of many statistically significant interrelations between what is heard and what is spoken, as well as what people like and vote for online.

But does this count as deliberative? Again, if our threshold is low, then yes: not only are the short-term dynamics interactive, responsive, and reflective of complex conceptual structures, but these short-term argument do plausibly seem to have long-term effects. And these long-term effects are at least as complex as the short-term effects appear to be. But one cause for worry is that, if one also desires, for instance, more policy-based speech and less emotion-based speech, the fact that both types tend to produce more of themselves means that it may be hard for the centrist, more emotional discussants to escape their domain without imposing undesired left-wing speech upon them. In particular, we might worry that speech too far from a user's "comfort zone" might naturally repel them rather than persuading. To test for this effect, we can include in the bivariate panel model a second independent variable, the difference between the quantity of the topic heard ($x_{it}$), and the listener's usual (and presumably preferred) quantity of topic X ($h_{it}$):[21]

$$y_{it} = \beta_1 x_{it} + \beta_2 (x_{it} - h_{it})^3 + f_i + \epsilon_{it}, \tag{3}$$

The most important dependent variable $y_{it}$ is the vote-based ideology, which presumably shapes what posts users seek out, read, and like, which in turn deeply affects all their other within-forum experiences. The result of estimating this model for ideology is that even more heard topics appear to have a significant effects than in the original bivariate model (1, 2, 3, 5, 7, 8), and all in the expected ideological direction, where more left-leaning topics produce more left-leaning preferences, and vice versa. More importantly, the cubic factor shows a very strong effect where the (purportedly) persuasive effects do steeply diminish with the difference between the heard topic and the user's usual preferences. This suggests that users cannot simply be bludgeoned out of one ideology preference and into another, and that there is a strong tendency to stick with what one knows.

But even so, there are many more subtle pathways from right to left and vice versa in

[21] The cube lets us detect if effects diminish with the distance between $x_{it}$ and $h_{it}$: when $\beta_1$ and $\beta_2$ have opposite signs, when $x_{it}$ is close to $h_{it}$, the effect of $x_{it}$ is simply $\beta_1$; when $x_{it}$ is far from $h_{it}$, the slope will diminish (in either direction), showing a reduced effect when the heard speech is too far from what the hearer prefers. Conversely, when $\beta_1$ and $\beta_2$ have the same signs, the effect will intensify with distance.

Figure 7, where a series of small steps can gradually draw a user from one side to the other, potentially accounting for the many users who do appear to change significantly over the 2009-2010 period.[22] This is presumably why users bother to argue at all – or at least, that subset of users who choose to engage in the more responsive exchanges characteristic of our model of deliberation as the exchange of networked ideas.

## 6 Conclusion

The purpose of the foregoing analysis has been to develop and test a theory of deliberation that differs from standard requirements of facticity, restrained emotion, and heavily engineered environments. This theory emphasizes one aspect of deliberation, but one that we consider to be at the core: the responsive, interactive process of two or more people thinking things through, just as individual deliberates. To find evidence of deliberation in the wild requires a complex model, with psychological and communicative elements, and a text-analytic component to systematically examine vast numbers of interpersonal interactions. Furthermore, a model of specifically *political* deliberation requires a domain that is political yet not utterly polarized, with measures of political ideology at least somewhat separate from the very text we wish to examine. As well as being substantively interesting in themselves, online political forums offer a unique set of political activists who are more engaged, knowledgable, and productive than all but a very few participants in other social media, or indeed most experiment subjects.

The examination of these two forums on the left and right reveals not just an excellent testing bed for a theory of deliberation, but also reveals a myriad of differences between the conceptual and deliberative structures on the left and right. Although the two sides do mirror each other in important ways (such as the centrist predilection for lauding great persons), in many other ways we see not a mirroring but a surprising continuum: the relatively right side of both

[22] For instance, looking again at the bivariate (left) panel in Figure 7, Topic 6 – the public option – seems to have a strong perverse effect, where hearing more harping on the public option causes a rightward shift in PCA ideology as well a decrease in left-wing topics like 1 (health insurance) and 10 (class). This is a plausible story given the vitriol on the left during the health-care debate, although an examination of the interaction effects between multiple topics, such as emotions and policy, must remain for later work.

forums tends to be more consistent in writing what they like and in liking more personal, meta and emotional topics, and seem to exhibit less densely interconnected individual networks which perforce are more dominated by a few specific topics. The two sides also differ to the degree to which their online arguments appear deliberative, in the responsive sense taken to be essential here: the Democratic users reveal a more cohesive forum where authors are less likely to beat their own drum, and more likely to engage in the deliberative exchange of networked ideas; while the Republican forum users are somewhat more likely to repeat their own concerns independent of interlocutor.

But that said, the more fundamental conclusion is that both sides reveal a deliberative structure and conceptual networks, and both show a surprising degree of responsive discussion, especially given the reputations of internet discussions. Moreover, these interactions might plausibly even have long term effects on some of the participants. True, many of these effects act to reinforce existing ideological preferences, and many users are resistant to persuasion by speech that is too unlike their preferences. But the network model reveals many possible pathways for persuasion, many of which may well have drawn users across the internal aisles of their parties. Overall, we see strong evidence for serious deliberation: users with strong differences arguing politics over sustained periods, responding to each other's points and raising related points of their own as they hash through the political debates of the era. While such users may be relatively uncommon even among the political elite, we might do well to study the circumstances that are conducive to these more deliberative exchanges; such circumstances may be easier to come by than many suppose.

# Appendix

The topic model employed here largely follows Blei, Ng & Jordan (2003) and Blei & Lafferty (2009). The basic structure begins with the original Latent Dirichlet allocation model of Blei, Ng & Jordan (2003), although that model posits a Dirichlet function generating the document topics (replace $\mu$ and $\Sigma$ in Figure 8 with $\alpha$), whereas we are interested in the covariance structure of the topics, $\Sigma$, as in Blei & Lafferty (2007).

The fundamental LDA process is well described in Blei & Lafferty (2009), but since the process here diverges in a number of places, it is worth brief recapitulation. Say there are D documents, N unique words, and K topics. We begin by summarizing the documents with the usual "bag of words," where $W_{d,n}$ = the proportion of document $d$ that is word $n$.[23] We assume that each document $d$ has a distribution over all K topics, $\theta_d$, and Blei & Lafferty (2007) assume that these topics distributions in turn have been drawn from a multivariate normal $\eta_d \sim \mathcal{N}(\mu, \Sigma)$ (after transforming from the simplex to Euclidean space, $\eta_d = f^{-1}(\theta_d)$). In generating a document, we assume $\eta_d$ is first drawn, then transformed to compositional form $\theta_d$, and then for each word slot in the document a topic is assigned, where each of these is a multinomial draw from $\theta_d$, creating a matrix $Z_{d,n} \sim \mathcal{M}(\theta_d)$. On the right side of Figure 8 we have the word distributions associated with each of the K topics, $\beta_k$. Finally, the document matrix $W_{d,n}$ is created from $Z_{d,n}$ and $\beta_k$ by taking multinomial draws of words from the appropriate word distribution $\beta_k$ depending on which topic has been asigned to a given slot, ie, $W_{d,n} \sim \mathcal{M}(\beta_{[k=Z_{d,n}]})$.

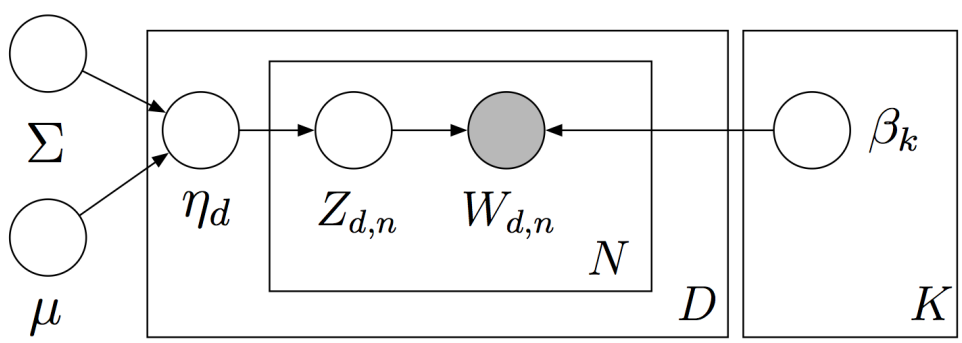


Figure 8: The generative model from Blei & Lafferty (2007).

[23] We keep the top 1000 more frequent words in the corpus, not including about 50 stop words such as "the", "of," etc. Because words with different stems mean different things ("American" ≠ "Americans"), we do not perform any stemming, nor do we perform any data transforms such as term frequency – inverse document frequency, although one can use tf-idf to select the stop words to exclude.

To summarize, we have:

D documents. N unique words. K topics.
Bag of words: $W_{d,n}$ = % of document $d$ that is word $n$.
Each document has distribution $\eta_d$ over $K$ topics.

| *Text* | *Topics* |
|---|---|
| Documents: $W = D \times N$ | $\eta = D \times K$ |
| Topic of word: $Z = D \times N$ | $\mu = K$ |
| Word dist per topic: $\beta = K \times N$ | $\Sigma = K \times K$ |

$$\eta_d \sim \mathcal{N}(\mu, \Sigma) \tag{4}$$
$$Z_{d,n} \sim \mathcal{M}(f(\eta_d)) \tag{5}$$
$$W_{d,n} \sim \mathcal{M}(\beta_{Z_{d,n}}) \tag{6}$$

At this point, the procedure here diverges somewhat from Blei & Lafferty (2007). Essentially, the LDA model from Blei, Ng & Jordan (2003) will be estimated, but with a change in the Dirichlet process on $\theta_d$. The posterior in Blei, Ng & Jordan (2003) is

$$p(\theta_d, z_{d,n}, \beta_k | w_{d,n}, \alpha) = \frac{p(\theta_d, z_d, \beta_k | w_d, \alpha)}{\int_\beta \int_\theta \sum_z p(\theta_d, z_d, \beta_k | w_d, \alpha)} \tag{7}$$

which is intractable due to the denominator, and is usually approximated using variational methods. The same approach is followed here, where we replace the original joint distribution with a product of parameter-specific distributions, and then maximize this product via coordinate ascent. In Blei & Lafferty (2007) even the variational approach is computationally challenging, fit via expectation maximization where the E step (the posterior distribution of the latent variables) is especially difficult due both to combinatorial explosion (of order $\mathrm{K}^N$) and the usual challenges of dealing with products of distributions that are not conjugate. To avoid these issues, we simply drop the requirement that $\eta_d$ be normally distributed, estimating $\mu$ and $\Sigma$ at the end of the process from the matrix of topics distributions per document.[24]

The other simplification here is that, unlike Blei, Ng & Jordan (2003), it is assumed all

[24]Since we are mainly engaged in descriptive analysis with these topic modes, and not (directly) in the estimation of substantively important quantities of interest, such simplifications have no more deleterious effect than all the other ones we make, such as beginning with the bag of words, or assuming various independences during variational procedures. The resultant topics are still quite meaningful, as are their correlations, and as we have seen, we are able to predict responses quite robustly as well, validating the empirical utility of this analytic simplification.

processes are multinomial (which is actually more like Blei & Lafferty (2007)), and thus the variational parameters are simply the same as the original parameters, with the exception of $\phi_{d,n,k}$, which replaces $Z_{d,n}$ by assuming that every word as a distribution over all topics, rather than a single topic.[25] Since Z and W are drawn from multinomials and we have replaced the Dirichlet process for $\theta$ with a simple multinomial, mean values are the maximal estimates for those parameters (conditional on the others), and $\phi$ is essentially just the (normalized) product of those two.

Essentially, the maximum is found very much as it is for the objective function in Blei, Ng & Jordan (2003):

$$\mathcal{L} = \sum_K \mathrm{E}[\log\, p(\beta_k|\eta)] + \sum_D \mathrm{E}[\log\, p(\theta_d|\alpha)] \tag{8}$$

$$+ \sum_D \sum_N \mathrm{E}[\log\, p(Z_{d,n}|\theta_d)] + \sum_D \sum_N \mathrm{E}[\log\, p(w_{d,n}|Z_{d,n}, \beta_k)] + H(q) \tag{9}$$

(Where all the conditional parameters are the variational replacements, and H is the entropy). The difference is that because we assume everything is multinomial we do not have the Dirichlet problems, we can retain the original parameter names (since nothing is changed), and given the parameters in (10), (11), (12), each is maximized conditional on the others by :

$$\beta_{k,n} = \frac{\Sigma_d W_{d,n}\phi_{d,n,k}}{\Sigma_{d,n} W_{d,n}\phi_{d,n,k}} \tag{10}$$

$$\theta_{d,k} = \frac{\Sigma_n W_{d,n}\phi_{d,n,k}}{\Sigma_{n,k} W_{d,n}\phi_{d,n,k}} \tag{11}$$

$$\phi_{d,n,k} = \frac{\theta_{d,k}\beta_{k,n}}{\Sigma_k \theta_{d,k}\beta_{k,n}} \tag{12}$$

Maxima are found by cycling through these conditional maximizations, (10) $\rightarrow$ (11) $\rightarrow$ (12) $\rightarrow$ (10) $\rightarrow$ ...etc. The first two are essentially taking word-weighted means, while the last is simply reconstructing $\phi$ and normalizing it. The result is a very fast maximization yielding topic distributions over documents and word distributions over topics. Finally, to calculate correlations

[25] Like Blei, Ng & Jordan (2003) we replace individual word indicators for every word with the normalized word frequency for that document, since every repeated word in the same document has the same topic distribution. This greatly speeds computation.

post hoc, the proportions $\theta$ must be transformed from the simplex to Euclidean space (using the centered log transform) to avoid the spurious correlations caused by working on the simplex.

In the semi-supervised approach employed here, we estimate the model on a subset of the most partisan users, as determined by the vote-based first principal component loadings. At this point, we retain $\beta_{k,n}$ and keep it unchanged, throw out the other parameters, and re-estimate them using the remainder of the users and the fixed parameters $\beta_{k,n}$. Additionally, because there are hundreds of thousands of comments in the sample, even with the efficiencies already introduced, estimating these parameters (particularly $\phi$) takes prohibitively long. But one additional efficiency can be found because, once $\beta_{k,n}$ is fixed, the effects of each document on each other is much diminished, and the model can thus be fit on subsets of the corpus. Doing this speeds the results greatly, and cross-validation suggests that document parameters are very similar regardless of being done in batches rather than as a whole, and regardless of the other members of the batch.

Finally, inferences about the multivariate normal topic distribution $\eta$ are made after maximization, without priors enforcing normalcy during the estimation process. This is important, because it shows that the predictive power of $\mathbf{P}$ is not something that derives from our prior assumptions, but which emerges naturally from the data.

The code for this topic model is available upon request, and an online version that quickly models uploaded documents can be found at nickbeauchamp.com.

## References


Adorno, T.W., E. Frenkel-Brunswik, D.J. Levinson & R.N. Sanford. 1950. "The authoritarian personality.".

Altemeyer, B. 1998. "The other "authoritarian personality"." *Advances in experimental social psychology* 30:47–92.

Alvarez, R.M. 1998. *Information and elections*. Univ of Michigan Pr.

Arellano, M. & O. Bover. 1995. "Another look at the instrumental variable estimation of error-components models." *Journal of econometrics* 68(1):29–51.

Arellano, M. & S. Bond. 1991. "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58(2):277–297.

Bartels, L.M. 1996. "Uninformed votes: Information effects in presidential elections." *American Journal of Political Science* pp. 194–230.

Beauchamp, Nicholas. 2011. "How to Scale Legislatures with Text: A comparison of methods, with applications to the US Congress and UK House of Commons." *Working paper* .

Blei, D.M., A.Y. Ng & M.I. Jordan. 2003. "Latent dirichlet allocation." *the Journal of machine Learning research* 3:993–1022.

Blei, D.M. & J.D. Lafferty. 2007. "A correlated topic model of science." *The Annals of Applied Statistics* pp. 17–35.

Blei, D.M. & J.D. Lafferty. 2009. "Topic models." *Text mining: classification, clustering, and applications* 10:71.

Druckman, J.N. 2001*a*. "On the limits of framing effects: who can frame?" *Journal of Politics* 63(4):1041–1066.

Druckman, J.N. 2001*b*. "The implications of framing effects for citizen competence." *Political Behavior* 23(3):225–256.

Druckman, J.N. 2004. "Political preference formation: Competition, deliberation, and the (ir) relevance of framing effects." *American Political Science Review* 98(4):671–686.

Farrar, C., J.S. Fishkin, D.P. Green, C. List, R.C. Luskin & E.L. Paluck. 2006. "Disaggregating deliberation's effects: An experiment within a deliberative poll." *Manuscript. New Haven, CT: Yale University* .

Ferejohn, J.A. & J.H. Kuklinski. 1990. *Information and democratic processes*. Univ of Illinois Pr.

Fishkin, James S, Robert C Luskin & Roger Jowell. 2000. "Deliberative polling and public consultation." *Parliamentary Affairs* 53(4):657–666.

Fishkin, J.S. 1991. *Democracy and deliberation: New directions for democratic reform*. Vol. 217 Cambridge Univ Press.

Franklin, C.H. 1991. "Eschewing obfuscation? Campaigns and the perception of US Senate incumbents." *The American Political Science Review* pp. 1193–1214.

Gelman, A. & G. King. 1993. "Why are American presidential election campaign polls so variable when votes are so predictable?" *British Journal of Political Science* 23(04):409–451.

Habermas, J. 1994. "Three normative models of democracy." *Constellations* 1(1):1–10.

Hamilton, J.D. 1994. *Time series analysis*. Vol. 2 Cambridge Univ Press.

Holbrook, T.M. 1999. "Political learning from presidential debates." *Political Behavior* 21(1):67–89.

Holtz-Eakin, D., W. Newey & H.S. Rosen. 1988. "Estimating vector autoregressions with panel data." *Econometrica: Journal of the Econometric Society* pp. 1371–1395.

Iyengar, S. 1994. *Is anyone responsible?: How television frames political issues*. University of Chicago Press.

Iyengar, S. & D.R. Kinder. 1987. *News that matters*. University of Chicago Press Chicago, IL.

Jacobs, L.R. & R.Y. Shapiro. 2000. *Politicians don't pander: Political manipulation and the loss of democratic responsiveness*. University of Chicago Press.

Krosnick, J.A. & D.R. Kinder. 1990. "Altering the foundations of support for the president through priming." *The American political science review* pp. 497–512.

Lasswell, H.D. 1948. "The structure and function of communication in society." *The communication of ideas* pp. 37–51.

Laver, M., K. Benoit & J. Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97(02):311–331.

Lodge, M., K.M. McGraw & P. Stroh. 1989. "An impression-driven model of candidate evaluation." *The American Political Science Review* pp. 399–419.

Love, I. & L. Zicchino. 2006. "Financial development and dynamic investment behavior: Evidence from panel VAR." *The Quarterly Review of Economics and Finance* 46(2):190–210.

Lupia, A. & M.D. McCubbins. 2000. *Elements of reason: Cognition, choice, and the bounds of rationality*. Cambridge Univ Pr.

Mansbridge, J., J. Hartz-Karp, M. Amengual & J. Gastil. 2006. "Norms of deliberation: An inductive study." *Journal of Public Deliberation* 2(1):7.

McCombs, M.E. & D.L. Shaw. 1993. "The evolution of agenda-setting research: Twenty-five years in the marketplace of ideas." *Journal of Communication* 43(2):58–67.

McCombs, M.E., D.L. Shaw & D.H. Weaver. 1997. *Communication and democracy: Exploring the intellectual frontiers in agenda-setting theory*. Lawrence Erlbaum.

Monroe, B.L. & K. Maeda. 2004. Talk's cheap: Text-based estimation of rhetorical ideal-points. In *annual meeting of the Society for Political Methodology*. pp. 29–31.

Mutz, D.C. 2006. *Hearing the other side: Deliberative versus participatory democracy.* Cambridge Univ Pr.

Mutz, D.C. 2008. "Is deliberative democracy a falsifiable theory?" *Annu. Rev. Polit. Sci.* 11:521–538.

Page, B.I. & R.Y. Shapiro. 1992. *The rational public: Fifty years of trends in Americans' policy preferences.* University of Chicago Press.

Pan, Z. & G.M. Kosicki. 1993. "Framing analysis: An approach to news discourse." *Political Communication* 10(1):55–75.

Poole, K.T. 2005. *Spatial models of parliamentary voting.* Cambridge Univ Pr.

Poole, K.T. & H. Rosenthal. 1985. "A spatial model for legislative roll call analysis." *American Journal of Political Science* 29(2):357–384.

Popkin, S.L. 1991. *The reasoning voter: Communication and persuasion in presidential campaigns.* University of Chicago Press.

Scheufele, D.A. 2000. "Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication." *Mass Communication & Society* 3(2-3):297–316.

Slapin, J.B. & S.O. Proksch. 2008. "A scaling model for estimating time-series party positions from texts." *American Journal of Political Science* 52(3):705–722.

Sniderman, P.M., R.A. Brody & P. Tetlock. 1993. *Reasoning and choice: Explorations in political psychology.* Cambridge Univ Pr.

Steenbergen, M.R., A. Bachtiger, M. Sporndli & J. Steiner. 2003. "Measuring political deliberation: a discourse quality index." *Comparative European Politics* 1(1):21–48.

Wilhelm, A.G. 2000. *Democracy in the digital age: Challenges to political life in cyberspace.* Psychology Press.

Zaller, J. 1992. *The nature and origins of mass opinion.* Cambridge Univ Pr.