Social News and Civic Engagement

Solomon Messing*

October 2013

Abstract

We conduct the first study on how socially-shared news-related content affects policy preferences and voter turnout. We use an AB test on Facebook launched during the 2012 election to instrument exposure to content and a large survey launched after election day to collect data on policy preferences and self-reported voter turnout. We first examine the correlation between socially-shared hard news exposure and civic engagement, with findings that are largely in line with the existing literature, then use an instrumental variables design to analyze the effect of seeing more socially shared political news on civic engagement. Our analysis shows that additional exposure to socially shared political content increased turnout and led to more consistent policy preferences among independents.

1 Introduction

The way in which public affairs information affects preferences and political behavior depends on the structure of the media. Scholars of media effects have argued that the relative homogeneity and popularity of television news coverage from 1950s to the early 1980s meant that the media could set the national agenda—the issues and concerns that affected political behavior and shaped elections (McCombs and Shaw 1972; Iyengar and Kinder 1987; Iyengar 1991). Of course, exposure to media today is certainly not uniform (see e.g., Zaller 1991, 1992; Sears and Freedman 1967); individuals have an overwhelming number of choices available, including a wide array of entertainment options, topically specialized publications, and openly partisan sources (Prior 2007; Bennett and Iyengar 2008). Some have argued that these developments mean the end of media effects on political behavior, as the majority of news consumers "opt-out" of news in favor of entertainment alternatives, while partisan consumers (whose preferences are much less malleable Zaller 1991; Price and Zaller 1993) habituate themselves to partisan news outlets (Mutz 2006; Iyengar et al. 2008; Iyengar and Hahn 2009).

*Jointly authored with Eytan Bakshy and Andrew Fiore. Thanks to Shanto Iyengar, Kimberly Gross, Kyle Dropp, Cliff Nass, Eytan Bakshy, Rebecca Weiss, Thomas Leeper, Nathaniel Swigger, Lilach Nir, Danny Hayes, Patricia Joseph, and other panel participants at the 2012 Annual Meeting of the Midwestern Political Science Association in Chicago, Illinois 2012. This research was generously supported by a Google Research Award and Stanford University. This draft was last revised on December 21, 2018

The last decade has made it clear that rather than simply tuning out, individuals are turning to social networks in order to allocate attention in an increasingly diverse and varied information environment. According to Pew, more than 60 percent of Americans have a Facebook account, and according to Nielsen, they spend more time on Facebook than any other website, by nearly two orders of magnitude (Nielsen 2011, 2012). About half of these users report exposure to news on the site (Mitchell et al. 2013). The social cues present in social media mean that individuals are more likely to attend to news shared in this medium (Messing and Westwood 2012), and others have suggested that citizens may be able to absorb the information necessary to participate effectively in the democratic process, via cues from media elites and social networks (Popkin 1994; Lupia and McCubbins 1998; Ryan 2011; Sokhey and McClurg 2012), suggesting that our social networks should affect political knowledge, preferences and behavior.

However, these factors likely depend on each other and interact dynamically over time—making it tremendously challenging to credibly answer the following question: How does information encountered in social networks affect opinion and behavior (e.g., Siegel 2013)? We utilize an exogenous source of variation in the amount of news-related content in user's News Feeds on Facebook, introduced via an A/B test, as an instrument to identify the effect of socially shared public affairs content on policy preferences and voter turnout. We first examine the correlation between hard news exposure and civic engagement, then use an instrumental variables design to analyze the effect of seeing more socially shared political news on civic engagement. Our analysis shows that additional exposure to socially shared political content induced more consistent policy preferences among independents and increased self-reported turnout.

Measure of Exposure to Shared Public Affairs Content

We first classify the news media content that friends of survey-participants shared from the beginning of 2012 to election day. During the 72-day study period, participants' friends shared 9.8 million URLs, 6.5 million of which were unique. These items appeared in participants' news feeds, which displayed a photograph, story title, domain, and a two to three sentence summary associated with the item, which are produced by the originating website. We classify the text of each of these summaries as public affairs or not, then determine which items appeared in each person's news feed using exposure-logging data.

Our classification strategy is to use the information embedded in URLs as labels indicative of public affairs content, then build a model to identify words and phrases predictive of content related to public affairs during the period under study. Once we have a model of these words and phrases, we can use it to classify unseen texts (an approach commonly known as supervised machine learning, see for example, Hillard et al. 2008; Hopkins and King 2010). We label as "public affairs content" all items that originated from our list of news domains *and* whose urls contained the tokens "politi," "usnews," "national," "election," "obama," and "romney." We then took content whose URLs contained variants of "sports,"

"entertainment," "arts," "style"[1] (from any URL) and label these items as "soft content."[2] This training corpus of labeled item summaries consists of 102,651 documents—77,866 soft items and 24,785 public affairs items. For each of these documents, we then computed unigram, bigram, and trigram features—counts of single words, word pairs, and word triplets, respectively.[3] Next, we remove features that occurred in more than half of all documents, or in only a single document, leaving 348,643 features. We represent each document (rows) and each feature (columns) in a matrix, X, which we represent in compressed sparse row (CSR) format.

Finally we fit a linear Support Vector Machine (SVM) classifier to the data, which finds the vector that maximizes separation between classes in multiple dimensions. We fit the model with a ridge penalty and hinge loss using the Python package "Scikit-Learn," (which calls the C++ "liblinear" library Pedregosa et al. 2011; Fan et al. 2008). The ridge penalty is desirable under conditions of collinearity, because it shrinks highly predictive coefficients toward the same value (compared to the situation in which one coefficient dominates the rest in unpenalized models). This is especially appropriate to deal with the problem of polysemy in text classification, wherein multiple words mean the same thing.

This approach to generating labels has several advantages over relying on conventional human-coded labels. First, the large number of documents we model helps to overcome some of the sparsity in our features that would otherwise lead to highly variable predictions. This sparsity is due to the rich variety of words used in human language. It is complicated by the curse of dimensionality—as the number of features increases, the sampling density is diluted by $N^{\frac{1}{p}}$, meaning the variance of our predictions will increase with each new feature. This is particularly problematic when attempting to classify public affairs content, in which the words used are even more highly variable and change based on current events.

Validation tests reveal that our classifier yields extremely accurate results. We assess the performance of our method using *cross-validation* (e.g., Hastie et al. 2009). Here we employ ten-fold cross-validation, which entails fitting a model using 9/10ths of the articles in our dataset, then assessing out-of-sample performance based on the remaining 1/10th. This is done ten times with each tenth of the articles, and the results are averaged. Based on ten-fold cross-validation, the classifier achieves accuracy of 0.946, precision of 0.933, and recall of 0.835.[4] This means that 95% of our out-of-sample classifications agreed with our labels; that given that a document is about public affairs, we identified it as such 84% of the time; and given that we made a prediction that a document was about public affairs, it was actually public affairs content 93% of the time. However, it is also desirable to examine randomly drawn examples of each document that our classifier predicts as about public affairs or not from the full corpus of content (e.g., Grimmer and Stewart 2013). Table 1 below presents

[1]The inclusion of "style" also captures sections like "lifestyle." URLs and matching expressions were all processed as lower-case.

[2]While one might suggest that this classification captures "hard" and "soft" news, we do not want to suggest we are distinguishing between, say, news about a political scandal, which many might consider "soft," and coverage of a bill being introduced in the House of Representatives, which is more clearly "hard news."

[3]We remove non-English items and pre-process each summary by removing any HTML or English stop words like "a," "the," "and," and so on.

[4]F1, the harmonic mean of precision and recall, is 0.881.

a sample of the text for both the public affairs and "soft" categories. As the classification metrics suggest, the classifier does extremely well, performing classifications as we would expect.[5]

[5]It should be noted that this classifier is suitable for formal media prose and is inappropriate for the much harder problem of classifying informal discourse.

Table 1: Random Sample of 10 Articles per Class

Class Label	Sample of Text Summary
Public Affairs	In a rare glimpse into his tenure at Bain Capital, Mitt Romney writes about lessons learned and how they prepare him for the presidency...
Public Affairs	If you believe that the United States should legalize dogfighting because we allow humans to fight, fear not. You've got an ally in the...
Public Affairs	You don't need to be born in the U.S. to be president?
Public Affairs	The killings heighten worries about how the coalition troops can protect themselves while working at close quarters with their Afghan...
Public Affairs	Vice President Biden's handlers appear to be going to new lengths to keep their boss's antics and open to interpretation quips from ...
Public Affairs	WARSAW—A Mitt Romney aide told reporters to "shove it" Tuesday morning after the American press corps here shouted questions ...
Public Affairs	If Obama isn't...
Public Affairs	As reported Thursday by Katrina Trinko of National Review. If you are on the Obama fundraising e-mail list, you may have gotten a ...
Public Affairs	Over the past couple weeks some of Barack Obama's top fans have openly deserted him, expressing a commonly heard frustration...
Public Affairs	China steps up campaign against Ramadan fasting for Uighurs; experts fear backlash...
Soft	Chad Johnson and his wife Evelyn Lozada got into an altercation over the weekend, which ended wi...
Soft	Check my route, see more data and leave a comment by clicking the link above..
Soft	Puzzled Hearts Collection...
Soft	FOX Sports content...
Soft	Watch the best free live streaming tv and radio Fresh Radio 102.9 FM Windhoek 90.1 FM Oshakati ...
Soft	First Congregational United Church of Christ, 4600 Hamilton Blvd, welcomed a new pastor, the Re...
Soft	Bob Knight's analogy between handling stress and being raped, expressed during a documentary on ...
Soft	Do you have nightmares about going to school? Do you hate going to school because you know you ...
Soft	With fullback tight end David Johnson lost for the season due to a torn ACL, the Steelers have a...

Empirical Context

Before presenting our causal estimates, we note the striking correlation between exposure to public affairs content on Facebook and our outcomes of interest, voter turnout, and policy preferences. Individuals who are exposed to more socially shared public affairs content are more likely to self-report that they had voted (Figure 1). Furthermore, partisans who are exposed to more news are more likely to report more consistent policy preferences. Nonpartisans who are exposed to more news also report more consistent preferences in line with those favored by Democrats (Figure 2). It's worth noting that these non-partisan identifiers are more likely to have liberal policy preferences in the first place—a recent Pew study found that more than half of individuals born after 1981 identify as Independents but vote heavily Democratic (Pew 2014).

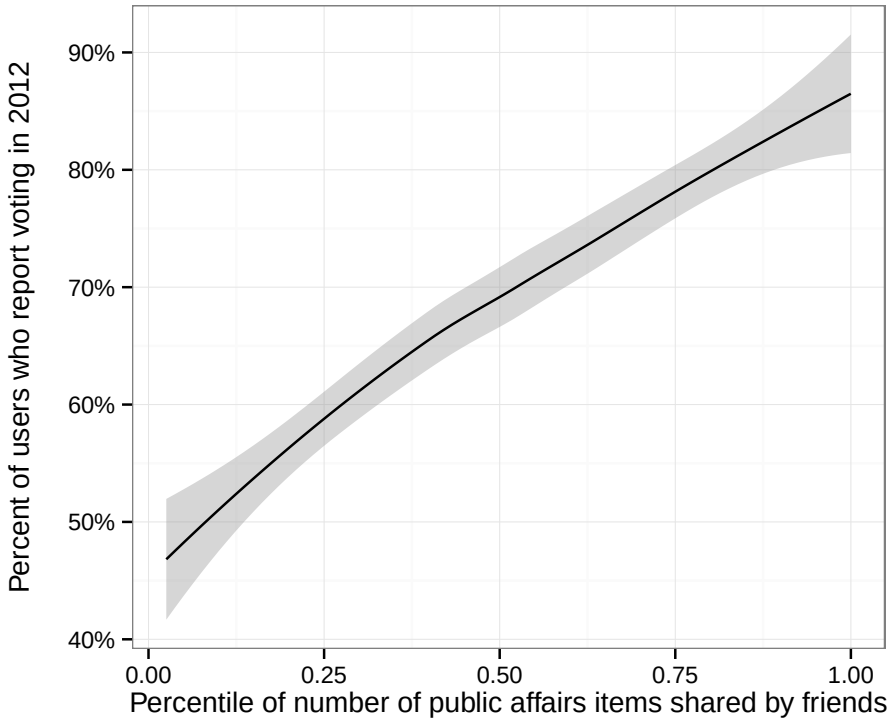

Figure 1: Individuals who tend to be exposed to more public affairs news stories have greater self-reported levels of turnout. Data is from individuals who saw the unaltered version of News Feed. Line fit to a loess curve, band indicates standard errors.

These observations are consistent with an array of prior research shows that consumption of public affairs media is associated with higher levels of political activism, and voter turnout, and more extreme policy preferences. However, it is difficult to determine whether media consumption itself can cause increases in these outcomes, or whether this relationship arises due to the influence of some third factor such as general political interest and knowledge. In the present study, exposure is driven primarily by what an individual's *friends* share, and not individual-level habits per se. Because individuals' attitudes, beliefs, and knowledge are correlated in friend networks due to homophily (Huckfeldt and Sprague 1995; Goel et al.

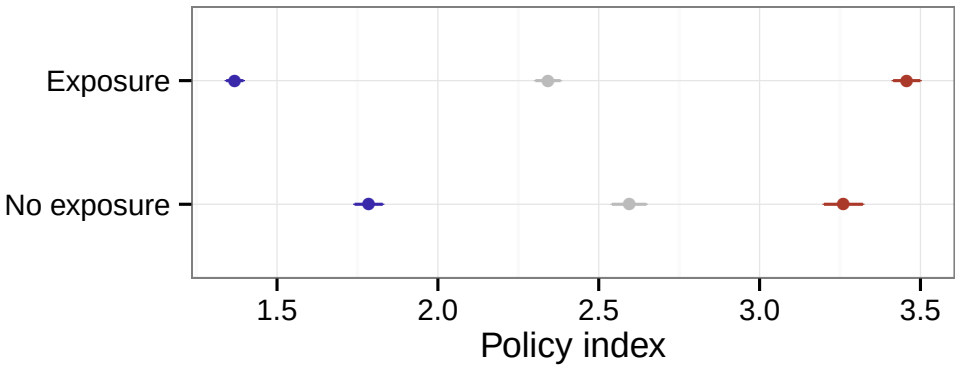

Figure 2: Observationally, individuals with more exposure to public affairs stories on Facebook hold more consistent policy preferences (higher index values correspond to policy preferences most consistently held by Republicans). Lines indicate standard errors.

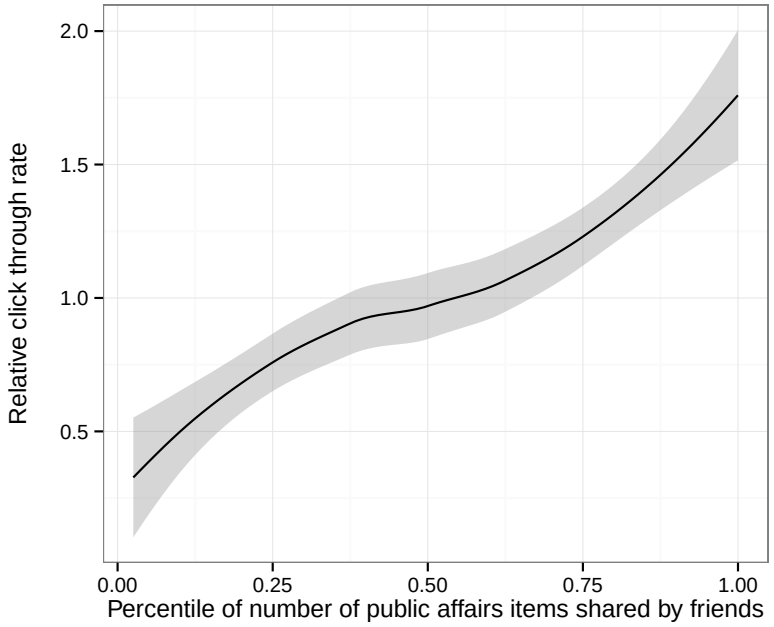

Figure 3: Individuals who have more friends sharing public affairs content are more interested in public affairs content, as indicated by click through rates (CTR). Plot shows CTR relative to public affairs stories' average CTR, as a function of the number of public affairs items shared by friends. The horizontal axes is percentile-transformed to show how CTR varies over the entire population. Data is from individuals who saw the unaltered version of News Feed. Band indicates standard error.

2010), we also expect that an individual's propensity to consume content is also correlated with how much content their friends share. This correlation is illustrated by Figure 3.

Finally, an individual's propensity to engage with public affairs content on Facebook is related to the highest "slot" in which it appears in the News Feed. Figure 4 shows that

individuals click on public affairs content that appears in the news feed in the first slot at nearly twice the rate as content that makes it only to the fourth slot. These order effects appear across a wide range of phenomena, from ballot order effects that can influence election outcomes (Koppell and Steen 2004) to lab-controlled news reading interfaces (Messing and Westwood 2012). Additionally, content is sorted based on a dynamic model that favors content with high interaction rates (among other factors), which may contribute to the observed relationship between click-through rate and position. Regardless, public affairs content that appears in the top five slots comprises many of the shared links that individuals click to access the content at the originating website.

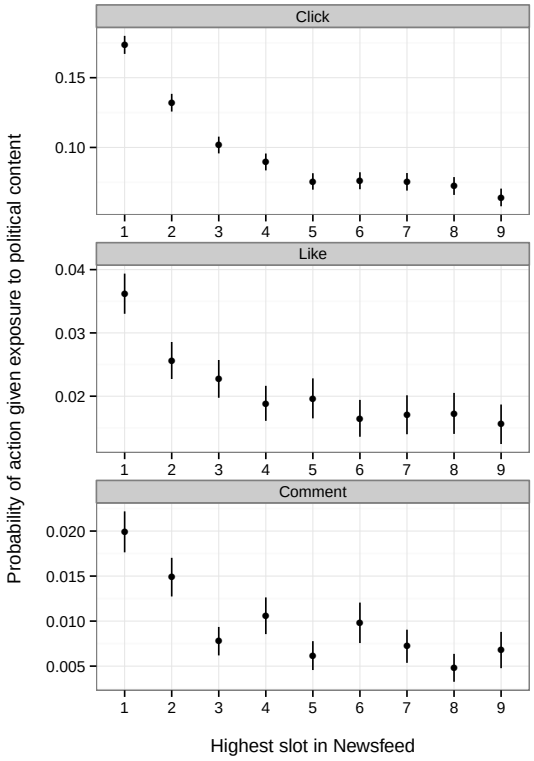

Figure 4: The likelihood of engaging with public affairs content is highest for content that appears in the top slot in an individual's newsfeed, and decreases as the highest slot reached declines. Lines indicate standard errors.

To examine the causal relationship between exposure to socially shared news on Facebook and these outcomes—turnout and policy preferences—we rely on an instrument that affected exposure to socially shared public affairs content as explained below.

Causal Analysis

2 The Instrument

Beginning on August 26, 2012, Facebook tested a version of the News Feed product that gave a slight boost to posts containing links to news media websites shared by friends when ordering content for display. 1.1 million U.S. Facebook users were randomly assigned to this version of News Feed.[6] These posts contained both links to public affairs news content and other content that appears on news websites, including sports and entertainment articles. Randomization was orthogonal to key pre-treatment demographic and behavioral measures (full randomization check provided in the Appendix). We compare these individuals with 560 thousand other Facebook users whose news feed remained unaffected by this or any other testing during the period in question.

3 Outcome Measures

In the days following the 2012 election (November 7 to November 19), 75 percent of these individuals over the age of 18 were asked to complete a survey that measured voter turnout, policy preferences, attitudes toward candidates and issues, political knowledge, and standard demographics (see Appendix for details). Individuals needed to login to Facebook to see the invitation. The response rate (among all invitees, irrespective of whether they logged in) was 1 percent, yielding 8.3 thousand participants assigned to the test version of News Feed and 4 thousand participants whose feed remained unaffected, who completed at least partial responses.[7] The survey included standard questions on policy preferences, media consumption, reports of voting behavior, and feelings toward political parties and officials. We provide the questionnaire and our analysis indicating a lack of differential non-response to our survey (which would also constitute differential attrition) in the Appendix.

In our analysis we focus on reported turnout and policy preferences. We measure turnout using the American National Election Study question wording. To capture policy preferences, we computed a six-point index of policy attitudes based on the extent to which respondents reported supporting the Patient Protection and Affordable Care Act ("Obamacare"), gun control, lower levels of federal spending, restricting access to abortion, and federal legalization of same sex marriage—respondents' answers to these questions were most strongly predictive of party affiliation (see appendix).

[6] Randomization relied on simple random assignment and persisted across users for the entire time period in question.

[7] Of these respondents, 48 percent answered at least 90 percent of the questions on the survey.

4 Results

4.1 Instrumented Exposure to Public Affairs Content

Having used the method described above to classify content that appeared in the news feed as public affairs content during the 72-day test period, we examine the number of items that made it to the top three slots in the News Feed for each set of respondents. Users who saw the test version that gave socially shared news a boost in the ordering in News Feed were exposed to 40 percent more public affairs content on average than those who saw the unmodified control version (Figure 5).

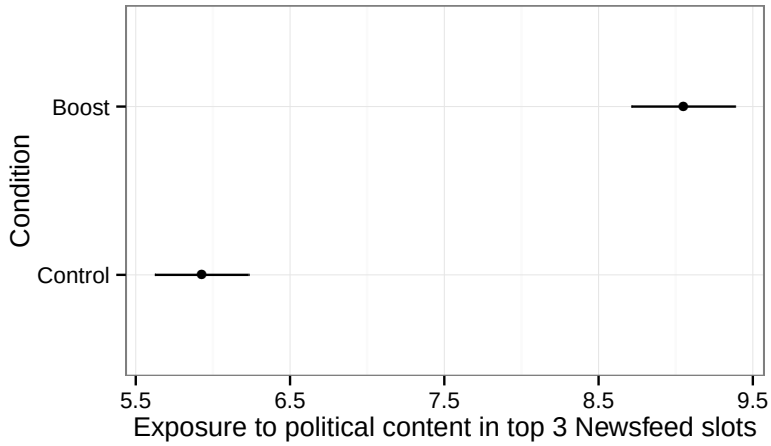

Figure 5: Those who saw the test version that gave socially shared news a boost in the ordering in News Feed were exposed to 40 percent more public affairs content. Lines are standard errors.

Because we have a source of randomized variation in exposure to public affairs content, but not a deterministic increase for all respondents, our design naturally lends itself to an instrumental variables modeling strategy (Sovey and Green 2011).[8] Our instrument—assignment to receiving the boosted levels of public affairs content in news feed (Z)—satisfies the exclusion restriction, because assignment is random, and only affects outcomes through its effect on increased exposure to public affairs news (D). And, as we will show below, it is correlated with the outcomes of interest, a relationship we refer to as the intent to treat (ITT) effect from here on.

We estimate these models using two stage least squares (2SLS), employing the standard correction in the second stage to calculate the covariance matrix correctly, as implemented in the R package AER.

4.2 Turnout and Political Salience

First, we compare self-reported turnout measure and our (post-treatment) political interest measure among those who saw the test version versus those who saw the unmodified version to establish that our instrument Z affected our outcome Y (turnout). A simple comparison of the proportion of respondents who report voting reveals a significant effect on self-reported turnout—70.9 percent of those who saw the test version of News Feed voted while 68.9 percent of those who saw the unmodified version report voting in the election (Wilcoxon rank sum test, $W = 10611494$, $p = 0.025$, two-sided). The test version significantly increased self-reported voter turnout.

[8]The exact number of news stories an individual was exposed to (D) may be correlated with their existing political preferences and voting tendencies (as illustrated in previous sections). However, that does not violate the assumptions of our instrumental variables setup.

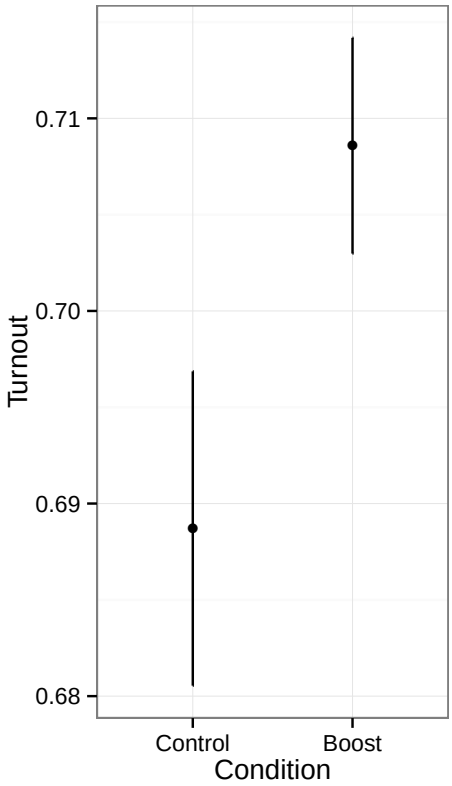

Figure 6: The ITT shows that the test version significantly increased self-reported voter turnout.

	Turnout	Political interest
(Intercept)	0.605^{***}	2.509^{***}
	(0.047)	(0.109)
Exposure (log_2)	0.055^{*}	$0.101^{\dagger}$
	(0.027)	(0.061)
N	9807	8922

Table 2: Two-stage least squares estimates of the effect on self-reported turnout and political interest. The instrumental variable, Z, is random assignment to the test version of News Feed with socially shared public affairs content more prominent in the ordering of content.

Our measure of political interest provides some evidence of why exogenous exposure to socially shared news increased turnout—suggesting the possibility that encountering public affairs information shared by one's peers raised the salience of politics. Indeed, individuals who saw the boosting test version of News Feed report slightly higher levels of political interest (more closely following governmental affairs) ($M = 2.69$) than those who saw the control version ($M = 2.66$, $W = 8565374$, $p = 0.097$). This is consistent with evidence from a natural experiment showing increased levels of political interest and turnout in response to exogenous media exposure from shared media markets in Switzerland (Butler and De La O 2011).

We now present our instrumental variables regression of voting on exposure to public affairs content. Because our exposure measure is heavily skewed, we transform it to a base 2 logarithmic scale. We also transform our political interest measure to a zero-one scale to make it easier to interpret. Our estimate of the effect for those who saw more public affairs news in the test group, or the local average treatment effect (LATE, Table 2), shows that each two-fold increase in the number of public affairs URLs to which users were exposed caused a 5.5 percentage point increase in the likelihood of voting. However, LATE effect of exposure on political interest is small and noisy.

4.3 Policy preferences

In addition to reporting higher levels of political interest and turnout, nonpartisans in the test group show more consistent policy preferences (a pattern in line with the literature on political knowledge and policy preferences, see e.g., Bartels 1996; Carpini and Keeter 1997). An ordinary least squares regression reveals an estimate of the ITT on the policy index among nonpartisans equivalent to roughly one-sixth the effect of Democratic party affiliation, or one-thirteenth the effect of Republican party affiliation (see also Figure 7, $\beta = -0.11, SE = 0.04, T = -2.48, P = 0.01$). Whether this arose from knowledge gain, argument

frequency, argument quality, opinion activation, or something else remains unclear.[9][10]

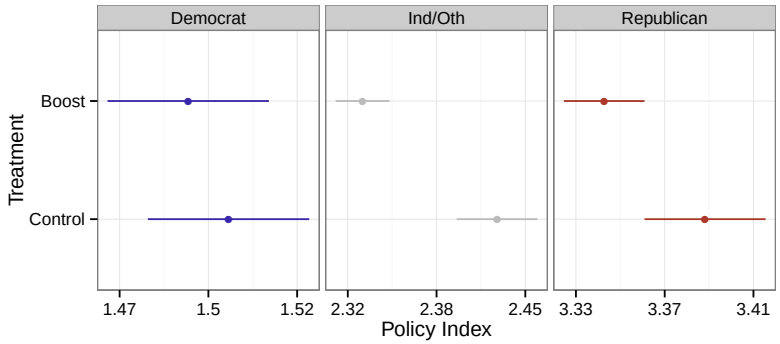

Figure 7: An examination of the ITT reveals more consistent policy preferences among nonpartisan identifiers in the test group.

[9]The main effect is similar in magnitude, though because it is driven by nonpartisans, we present only the conditional effect for this group. Of course, because political party identification explains so much of the variation in policy preferences, it is unproductive to estimate this effect without the precision gained by conditioning on these variables.

[10]The effect of encountering additional socially shared news was strongest on social issues (abortion and gay marriage), but persisted when adding other policies to the index, including federal spending levels, health care, and gun control. We use the broader index in the above analysis.

	All	Independents	Democrats	Republicans
(Intercept)	2.535***	2.773***	1.519***	3.632***
	(0.101)	(0.200)	(0.138)	(0.259)
Exposure (log_2)	−0.128*	−0.263*	−0.013	−0.176
	(0.056)	(0.112)	(0.068)	(0.163)
Democrat	−0.783***			
	(0.027)			
Republican	1.022***			
	(0.029)			
N	6873	2194	2854	1825

Table 3: Two-stage least squares estimates of the effect on policy preferences. The policy index ranges from zero to five, where lower values indicate more consistent support for policies favored by Democrats and higher values indicate more consistent support for policies favored by Republicans. The instrumental variable, Z, is random assignment to the test group that saw additional public affairs content.

Our instrumental variables analysis reveals the LATE in terms of exposure to public affairs content. Again, our exposure measure is heavily skewed, so we transform it to a base two logarithmic scale. We also include indicators for whether the individual identifies as a Democrat or Republican as covariates in both stages of the model. As shown in Table 3, a two-fold increase in the number of public affairs urls to which nonpartisan users were exposed resulted in a -0.26 point shift along the index.

One possibility that explains these results is that respondents who saw more public affairs content provided responses that are more internally consistent, by virtue of being better informed about the issues. That possibility is consistent with the literature on public affairs information, which generally finds that informed individuals are more likely to hold more consistent and stronger policy preferences (Bartels 1996, 1988; Zaller 1992; Krosnick and Brannon 1993; Carpini and Keeter 1997) and is compatible with work showing that individuals who do not have attitudes on an issue often form them in the process of exposure to particular arguments and ways of framing different policies (see Druckman et al. 2012). This literature suggests that individuals in the treatment condition who did not identify as Democrats or Republicans could have formed more consistent policy preferences in the process of being exposed to frames in line with those favored by Democrats. Indeed, nonpartisans were exposed to substantially more content that was shared by individuals identifying as liberal in their Facebook profiles than those identifying as conservative (see Appendix). It's also possible that individuals simply formed more left-leaning preferences in response to public affairs information (e.g., Vallone et al. 1985, though we did not detect an effect for partisans), an interpretation that could explain a similar phenomenon found in Gerber et al. (2009).[11]

[11] Gerber et al. (2009) found higher support for Democratic candidates, though not policy preferences, among individuals who received either the Washington Post or the right-leaning Washington Times during a 10-week field experiment on media exposure. We did not find a significant ITT or LATE effect on af-

Discussion

We provide preliminary evidence that over the course of a national election, exposure to higher levels of socially shared public affairs content increased self-reported turnout, while generally leading nonpartisans to report more consistent policy preferences. We observed limited evidence for a positive effect on political knowledge. We did not observe a significant effect on issue salience, political knowledge, belief in falsehoods, feelings toward Mitt Romney or Barack Obama.

The encouragement design used here is especially appropriate to document the effects of real-world exposure. By subtly increasing exposure to socially shared media content for individuals randomly assigned to the test group, this design ensures internal validity while maintaining external and ecological validity. This is the first study of which we are aware to utilize a source of random variation in exposure to public affairs content—socially shared or not—for such a long period of time, during an election season, and in which modern survey instruments are available to assess relevant outcomes.[12][13] Yet, by obtaining precise measures of exposure to public affairs content, we avoid the pitfalls of relying on self-reported measures of exposure to public affairs information (thought to be plagued by issues related to faulty recall and social desirability bias, see Price and Zaller 1993; Prior 2009; Vavreck 2007).

Nonetheless, the evidence here should be taken as highly preliminary. In light of the number of respondents, the effects here are noisy and the actual effect size may be much smaller than the ITT effects or LATE that we present here. In addition, the local average treatment effects we document only identify the effect of increasing exposure to news on Facebook shared by peers for those who could be exposed to more news (Sovey and Green 2011). Some individual's friends simply do not share a sufficient amount of public affairs content such that additional news could be shown, and, as we find in the empirical context section above, this group has a lower propensity to vote and holds less extreme policy preferences.Of course, as discussed above, we also cannot shed much light on the mechanism behind turning out to vote, nor non-partisan identifiers expressing more consistent policy preferences. It's important to note that we tested for but did not detect a significant ITT effect on candidate affect measured via a feeling thermometer, a political knowledge battery, and belief in falsehoods battery.

Nevertheless, this work helps address long-standing questions about how increases in civic engagement among individuals' peer groups can influence their opinions and behavior (Berelson et al. 1954; Katz and Lazarsfeld 1955; Delli Carpini 2000; Shah et al. 2001; Bennet 2008; Gerber et al. 2008; Pasek et al. 2009; Gil de Zúñiga et al. 2012), and about how social

fect measured via a feeling thermometer or self-reported vote for Obama, though both were positive and approached significance.

[12] Gerber et al. (2009) found marginally higher support for Democratic candidates, though not policy preferences nor turnout, among individuals who received either the Washington Post or the right-leaning Washington Times during a 10-week field experiment on media exposure. We did not find a significant ITT or LATE effect on candidate affect measured via a feeling thermometer or self-reported vote choice, though both were in the same direction as Gerber et al. (2009).

[13] It should be noted that an examination of how increased exposure to socially shared public affairs content affects political knowledge proved inconclusive. (An OLS estimate of the effect of the treatment on a knowledge index, which included an interaction with an indicator for everyday Facebook use, was not significant, $\beta = 0.01, SE = 0.01, T = 1.065, P = 0.287$.)

factors and algorithmic aspects of media distribution shape our information environment (Lewin 1947; Lazarsfeld et al. 1948; Schramm 1949; Freedman and Sears 1965; Pariser 2011; Bakshy et al. 2012; Muchnik et al. 2013).

These findings underscore the wide variance in exposure to public affairs content in social media, which depend on the behavior of one's social contacts. Of course, this phenomenon is not new (having been discussed at length in, for example, Sears and Freedman 1967). Yet, the massive increase in horizontal channels of distribution (i.e., social media and email) in recent years suggests this dimension of information distribution is more important than in previous times. Our work shows some of the consequences of this kind of exposure to public affairs content for political preferences and behavior. The shift in media distribution and consumption in favor of social channels means that this study constitutes an important step in understanding the implications for political preferences and behavior. This shift also means that this study should, if anything, grow in terms of relevance and external and ecological validity as the prominence of social media for these purposes continues to rise.

The fact that these effects depend so heavily on exposure to what our friends share means that factors related to social influence—tie strength, status, and perceived expertise—may interact with content to drive attention and ultimately influence outcomes relevant to politics, (hypotheses suggested by our exploratory analyses of heterogeneous treatment effects in this data, see SI). Indeed, this work supports the hypothesis that exposure to elite discourse is increasingly dependent on endogenous social factors (Bennett and Iyengar 2008), and suggests that as the prevalence of socially shared media increases, agendas should be expected to grow increasingly varied and individualized. Future research should explore how these social factors impact exposure and interact with content to affect political preferences and behavior.

5 Appendix

5.1 Randomization check

Randomization was orthogonal to friends' propensity to share public affairs content, as well as political affiliation and other relevant demographics. We take self-reported political affiliation and scale it from -1 (Liberal/Democrat) to 1 (Conservative/Republican, full details available in the Supplementary Information), then test for differences between the individuals who saw the test News Feed and others who saw the unmodified feed using the Wilcoxon Rank Sum test. Differences between those who saw the test version ($\mu = .017$) and control version ($\mu = .016$) were not significant ($W = 891690336$, $p = .47$, two-sided). Likewise, the proportion of males to females did not differ significantly for individuals who saw the test version ($\mu = .427$) versus control version ($\mu = .427$), ($W = 63876961992$, $p = .58$, two-sided). In addition, any difference in age between individuals who saw the test version ($\mu = 34.9$) and control version ($\mu = 34.9$) was not significant ($T(498641.9) = 0.46, p = .64$, two-sided).

5.2 Non-response/attrition

Because we only observe outcomes for individuals who responded to the survey, one threat to validity is differential response: that individuals who fill out the survey might differ between those who saw the test version and control version of News Feed.

We designed our survey recruitment and question wording to minimize the salience of politics at the time of deciding to respond. Respondents were initially recruited into the survey with the prompt "We'd like to hear from you. Please take 2-3 minutes to share your thoughts and opinions." Furthermore, our measures of turnout and political preferences using ANES wording were issued after a standard battery of demographic questions.

Because so many Facebook users self-report their political affiliation and demographics, our study allows for an extensive assessment of survey non-response by subgroup. We model response rates as a binary outcome and condition on various demographic to examine the predictors of survey response.

Table 4 shows that those in the test condition are not significantly more likely to respond to the survey. Those who respond to the survey are substantially more likely to use Facebook, to report their political affiliation, to be embedded in networks that report political affiliation, and skew slightly left. However, we do not observe differential non-response in any of those subgroups. A likelihood ratio test between model specifications containing these subgroups (2) and including the treatment indicator and all treatment interactions (3), suggests that

(3) does not yield significantly more explanatory power ($\chi^2(11) = 8.74$, P = 0.35).[14]

[14]Treating age as an integer variable does not change these results.

Table 4: Predictors of survey response, logistic regression coefficients. Individuals who were more active on Facebook and individuals who were more involved in politics were more likely to respond to the survey. Results do not indicate differential non-response by test group.

	(1)	(2)	(3)
(Intercept)	-5.006^{***}	-5.489^{***}	-5.505^{***}
	(0.016)	(0.022)	(0.039)
Test	0.033		0.024
	(0.019)		(0.047)
Age 25-34		-0.247^{***}	-0.258^{***}
		(0.024)	(0.041)
Age 35-44		-0.211^{***}	-0.248^{***}
		(0.027)	(0.047)
Age 45-54		-0.104^{***}	-0.141^{**}
		(0.030)	(0.053)
Age 55-64		-0.234^{***}	-0.250^{***}
		(0.039)	(0.068)
Age 65 +		-0.331^{***}	-0.450^{***}
		(0.050)	(0.091)
Male		-0.188^{***}	-0.191^{***}
		(0.019)	(0.033)
Everyday FB user		1.640^{***}	1.660^{***}
		(0.019)	(0.034)
Conservative Affil.		0.632^{***}	0.587^{***}
		(0.036)	(0.064)
Liberal Affil.		0.776^{***}	0.858^{***}
		(0.033)	(0.056)
T × Age 25-34			0.016
			(0.050)
T × Age 35-44			0.055
			(0.057)
T × Age 45-54			0.055
			(0.064)
T × Age 55-64			0.025
			(0.083)
T × Age 65 +			0.173

			(0.109)
T × Male			0.006
			(0.040)
T × Everyday FB user			−0.029
			(0.041)
T × Conservative Affil.			0.066
			(0.077)
T × Liberal Affil.			−0.123
			(0.070)
Log-likelihood	−74540.236	−68871.163	−68865.607
Deviance	149080.472	137742.325	137731.215
AIC	149084.472	137762.325	137771.215
BIC	149109.311	137886.163	138018.889
N	1829907	1765128	1765128

5.3 Exposure to Content by Sharer's Partisanship

Above, we show that individuals who saw the test version of News Feed who do not affiliate with one group of partisans reported stronger and more consistent support for policies favored by Democrats than nonpartisans who saw the unaltered version, and we speculated in the discussion that this could be due to exposure to content that was framed in a particular way as they formed opinions on issues. One indication of partisan framing is whether an individual who shares an article identifies as a partisan herself. After all, sharing constitutes an endorsement (of the importance of the piece if not the way facts and opinions are presented therein). In Figure 8, we present the average proportion of public affairs items appearing in the top 3 slots in participants' newsfeeds by the ideology of the sharing friends. Nonpartisans tended to be exposed to a higher proportion of content shared individuals with a left-leaning ideology than a right-leaning ideology, consistent with the framing-opinion formation explanation.

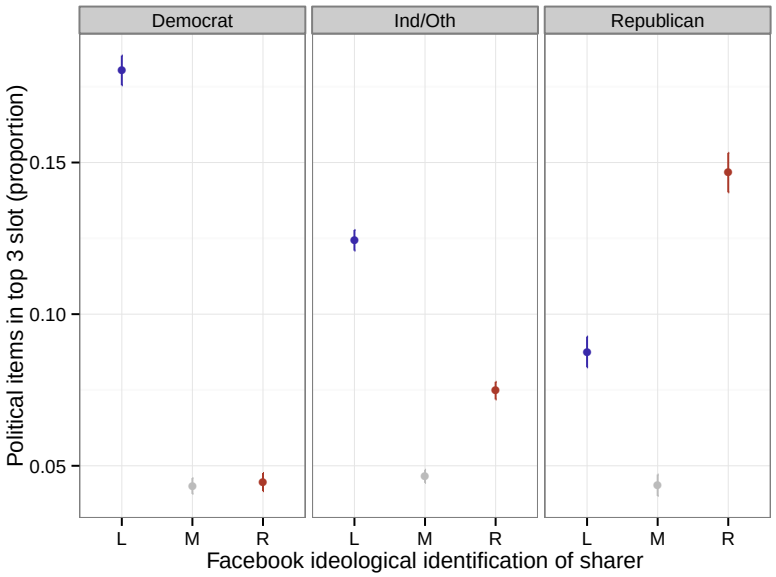

Figure 8: Nonpartisans were exposed to a higher proportion of content shared individuals with a left-leaning reported ideology than a right-leaning reported ideology. Content appearing in participants' newsfeeds by ideology of friends sharing. 3-point left, middle, right ideological categories based on a mapping from the top 500 self-reported "politics" designations from users profiles.

References

Bakshy, Eytan, Rosenn, Itamar, Marlow, Cameron, and Adamic, Lada (2012). The Role of Social Networks in Information Diffusion. *Proceedings of the 21st international conference on World Wide Web Pages* 1201.4145.

Bartels, Larry M. (1988). *Presidential Primaries and the Dynamics of Public Choice*. Princeton, N.J: Princeton University Press. ISBN 9780691022833.

Bartels, Larry M. (1996). Uninformed Votes: Information Effects in Presidential Elections. *American Journal of Political Science* 40(1):194–230.

Bennet, W. Lance (2008). Changing Citizenship in the Digital Age. In *Civic Life Online: Learning How Digital Media Can Engage Youth*, edited by W. Lance Bennet, 1–24. Cambridge, MA: MIT Press.

Bennett, W. Lance and Iyengar, Shanto (2008). A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication* 58(4):707–731.

Berelson, Bernard R., Lazarsfeld, Paul F., and McPhee, William N. (1954). *Voting: A Study of Opinion Formation in a Presidential Campaign*. Chicago: University Of Chicago Press. ISBN 0226043509.

Butler, Daniel M. and De La O, Ana L. (2011). The Causal Effect of Media-Driven Political Interest on Political Attitudes and Behavior. *Quarterly Journal of Political Science* 5(4):321–337. ISSN 15540634. doi:10.1561/100.00010041.

Carpini, Michael X. Delli and Keeter, Scott (1997). *What Americans Know about Politics and Why It Matters*. New Haven: Yale University Press. ISBN 9780300072754.

Delli Carpini, Michael X (2000). Gen.com: Youth, Civic Engagement, and the New Information Environment. *Political Communication* 17:341–349.

Druckman, James N., Fein, Jordan, and Leeper, Thomas J. (2012). A Source of Bias in Public Opinion Stability. *American Political Science Review* 106(02):430–454. doi:10.1017/S0003055412000123.

Fan, Rong-En, Chang, Kai-Wei, Hsieh, Cho-Jui, Wang, Xiang-Rui, and Lin, Chih-Jen (2008). LIBLINEAR: A Library for Large Linear Classification. *J. Mach. Learn. Res.* 9:1871–1874. ISSN 1532-4435.

Freedman, J. L. and Sears, D. O. (1965). Selective exposure. In *Advances in Experimental Social Psychology*, edited by L. Berkowitz, volume 2, 58–98. New York: Academic Press.

Gerber, Alan S., Green, Donald P., and Larimer, Christopher W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *The American Political Science Review* 102(1):33–48. ISSN 0003-0554. doi:10.2307/27644496.

Gerber, Alan S., Karlan, Dean, and Bergan, Daniel E. (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. *American Economic Journal: Applied Economics* 1(2):35–52.

Gil de Zúñiga, Homero, Jung, Nakwon, and Valenzuela, Sebastián (2012). Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication* 17(3):319–336. ISSN 1083-6101. doi:10.1111/j.1083-6101.2012.01574.x.

Goel, Sharad, Mason, Winter, and Watts, Duncan J. (2010). Real and Perceived Attitude Agreement in Social Networks. *Journal of Personality and Social Psychology* 99(4):611–621. ISSN 1939-1315. doi:10.1037/a0020697.

Grimmer, Justin and Stewart, Brandon M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* mps028. ISSN 1047-1987, 1476-4989. doi:10.1093/pan/mps028.

Hastie, Trevor, Tibshirani, Robert, and Friedman, Jerome (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*. New York: Springer, 2nd edition. ISBN 0387848576.

Hillard, Dustin, Purpura, Stephen, and Wilkerson, John (2008). Computer-Assisted Topic Classification for Mixed-Methods Social Science Research. *Journal of Information Technology & Politics* 4(4):31–46.

Hopkins, Daniel J. and King, Gary (2010). A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science* 54(1):229–247. ISSN 1540-5907. doi:10.1111/j.1540-5907.2009.00428.x.

Huckfeldt, R. Robert and Sprague, John D (1995). *Citizens, Politics, and Social Communication: Information and Influence in an Election Campaign*. Cambridge, UK: Cambridge University Press. ISBN 0521452988 9780521452984.

Iyengar, Shanto (1991). *Is Anyone Responsible?: How Television Frames Political Issues*. University Of Chicago Press. ISBN 0226388557.

Iyengar, Shanto, Hahn, K. S, Krosnick, J. A, and Walker, J. (2008). Selective exposure to campaign communication: The role of anticipated agreement and issue public membership. *The Journal of Politics* 70(01):186–200.

Iyengar, Shanto and Hahn, Kyu S. (2009). Red Media, Blue media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication* 59(1):19–39.

Iyengar, Shanto and Kinder, Donald R. (1987). *News that Matters: Agenda-Setting and Priming in a Television Age*. Chicago: University of Chicago Press.

Katz, Elihu and Lazarsfeld, Paul F. (1955). *Personal Influence: Social Forces*. New York: The Free Press.

Koppell, Jonathan GS and Steen, Jennifer A. (2004). The Effects of Ballot Position on Election Outcomes. *Journal of Politics* 66(1):267–281. ISSN 1468-2508. doi:10.1046/j.1468-2508.2004.00151.x.

Krosnick, Jon A. and Brannon, Laura A. (1993). The Impact of the Gulf War on the Ingredients of Presidential Evaluations: Multidimensional Effects of Political Involvement. *American Political Science Review* 87(04):963–975. doi:10.2307/2938828.

Lazarsfeld, Paul Felix, Berelson, Bernard, and Gaudet, Hazel (1948). *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lewin, Kurt (1947). Frontiers in Group Dynamics II. Channels of Group Life; Social Planning and Action Research. *Human Relations* 1(2):143–153. ISSN 0018-7267, 1741-282X. doi:10.1177/001872674700100201.

Lupia, Arthur and McCubbins, Mathew D (1998). *The Democratic Dilemma: Can Citizens Learn What They Really Need to Know?* Cambridge, U.K.; New York: Cambridge University Press. ISBN 0521584485 9780521584487 0521585937 9780521585934.

McCombs, Maxwell E. and Shaw, Donald L. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly* 36(2):176–187. ISSN 0033-362X. doi:10.2307/2747787.

Messing, Solomon and Westwood, Sean J. (2012). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation when Selecting Online News Media. *Communication Research* .

Mitchell, Amy, Kiley, Jocelyn, Gottfried, Jeffrey, and Guskin, Emily (2013). The Role of News on Facebook. Technical report, Pew Research Center.

Muchnik, Lev, Aral, Sinan, and Taylor, Sean J. (2013). Social Influence Bias: A Randomized Experiment. *Science* 341(6146):647–651. ISSN 0036-8075, 1095-9203. doi:10.1126/science.1240466.

Mutz, Diana C (2006). *Hearing the Other Side: Deliberative Versus Participatory Democracy*. New York: Cambridge University Press.

Nielsen (2011). State of the Media: Social Media Report Q3.

Nielsen (2012). State of the Media: Social Media Report 2012.

Pariser, Eli (2011). *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin Press. ISBN 1594203008.

Pasek, Josh, More, Eian, and Romer, Dan (2009). Realizing the social Internet? Offline social networking meets online civic engagement. *Journal of Information Technology and Politics* .

Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, Vanderplas, Jake, Passos, Alexandre, Cournapeau, David, Brucher, Matthieu, Perrot, Matthieu, and Duchesnay, Édouard (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Popkin, Samuel L. (1994). *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University Of Chicago Press, 1 edition. ISBN 0226675459.

Price, Vincent and Zaller, John (1993). Who Gets the News? Alternative Measures of News Reception and Their Implications for Research. *Public Opinion Quarterly* 57(2):133–164. ISSN 0033-362X, 1537-5331. doi:10.1086/269363.

Prior, Markus (2007). *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press, 1st edition. ISBN 0521675332.

Prior, Markus (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly* nfp002. ISSN 0033-362X, 1537-5331. doi:10.1093/poq/nfp002.

Ryan, John Barry (2011). Social Networks as a Shortcut to Correct Voting. *American Journal of Political Science* 55(4):753–766. ISSN 1540-5907. doi:10.1111/j.1540-5907.2011.00528.x.

Schramm, W. (1949). The Gatekeeper: A Memorandum. In *Mass Communications*, 175–177. Urbana, IL: University of Illinois Press.

Sears, David O. and Freedman, Jonathan L. (1967). Selective Exposure to Information: A Critical Review. *Public Opinion Quarterly* 31(2):194.

Shah, Dhavan V, Kwak, Nojin, and Holbert, R. Lance (2001). "Connecting" and "Disconnecting" With Civic Life: Patterns of Internet Use and the Production of Social Capital. *Political Communication* 18:141–162.

Siegel, David A. (2013). Social Networks and the Mass Media. *American Political Science Review* 107(04):786–805. doi:10.1017/S0003055413000452.

Sokhey, Anand Edward and McClurg, Scott D. (2012). Social Networks and Correct Voting. *The Journal of Politics* 74(03):751–764. doi:10.1017/S0022381612000461.

Sovey, Allison J. and Green, Donald P. (2011). Instrumental Variables Estimation in Political Science: A Readers' Guide. *American Journal of Political Science* 55(1):188–200. ISSN 1540-5907. doi:10.1111/j.1540-5907.2010.00477.x.

Vallone, Robert P., Ross, Lee, and Lepper, Mark R. (1985). The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre. *Journal of Personality and Social Psychology* 49(3):577–585.

Vavreck, Lynn (2007). The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports. *Quarterly Journal of Political Science* 2:287–305.

Zaller, John (1991). Information, Values, and Opinion. *The American Political Science Review* 85(4):1215–1237. ISSN 0003-0554. doi:10.2307/1963943.

Zaller, John R. (1992). *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press. ISBN 0521407869.