The Impact of Social Media Use on Voter Knowledge and Behavior in the 2015 UK Election: Evidence from a Panel Survey

Kevin Munger*, Patrick Egan
Jonathan Nagler, Jonathan Ronen, Joshua A. Tucker†

September 3, 2015

Abstract

The positive relationship between social media use and political sophistication is well established, but the causal connection is not. Using a panel survey that spans the 2015 UK election of respondents for whom we have a list of all tweets that appeared in their timelines during that period, we demonstrate that being exposed to tweets sent by media sources improves factual knowledge of political issues, but that the effects of being exposed to tweets sent by politicians are mixed. We also find weak support for the theory that being exposed to a more ideologically diverse set of tweets improves political knowledge.

*Corresponding author: km2713@nyu.edu

†Munger, Egan, Nagler and Tucker: New York University, Department of Politics and Social Media and Political Participation (SMaPP) lab; Ronen: New York University, SMaPP lab. This paper is a result of research collaboration between the SMaPP lab and YouGov, and we are extremely grateful to YouGov for its support of this project. The research of SMaPP lab is generously supported by INSPIRE program of the National Science Foundation (Award SES-1248077), the New York University Global Institute for Advanced Study, and Dean Thomas Carews Research Investment Fund at New York University. All of the authors contributed to theorizing, research design, and writing of the manuscript. Munger and Ronen wrote the computer programs, and Munger analyzed the data and prepared the tables, figures, and supplementary materials. We thank Duncan Penfold-Brown and Yvan Scher for additional programming and data science support.

1 Introduction

Twitter and other social media have become essential tools for the practice of modern politics. Politicians and parties tweet their views and can respond to constituents, and media outlets and journalists can spread their stories directly to their readers. However, the study of the effect of social media on individual political behavior is not fully developed. In particular, we still know very little about whether or not social media usage causes people to become more informed about politics, if it has no effect upon political knowledge, or, perhaps most intriguingly but also most worryingly, if it is associated with more *incorrect* information about politics? Does Twitter represent the democratization of discourse and an end to the stranglehold of a few elites on political information, or is it an echo chamber where partisan zealots take biased information and groupthink it further from the truth (Barberá et al., 2015)?

In this paper, we test a series of hypotheses related to the ways in which social media usage affects peoples' perceptions of two different types of political knowledge: answers to factual political questions, and being able to place parties' positions on issues correctly. These types of hypotheses, however are notoriously difficult to answer using cross-sectional data. It has been frequently observed that social media users are more politically informed, but the causal connection is murky: social media users also tend to be richer and better educated, two characteristics each associated with political knowledge (Carpini and Keeter, 1997). Furthermore, people looking for political information on social media may be people looking for political information elsewhere.

To address these concerns, we use innovative new data in our research design. First, our primary analysis takes place using a 4-wave panel survey of citizens of the UK conducted before, during and after the 2015 election campaign. This allows us to measure how individual levels of political knowledge – defined here as the ability to answer factual questions about political topics and to identify the positions of major parties on major issues – changed over time. This allows us to control for all unvarying individual-level covariates – including education, wealth,

political interest, etc. – and simply observe the way that people become more (or less) politically knowledgeable during the course of the campaign. However, even panel surveys ultimately depend on self-reported answers to measure social media usage and/or the content to which one is exposed on social media. Therefore, in order to establish an objective measure of exposure to political information, we leverage our access to the actual set of tweets to which our survey respondents could have been exposed to measure both the issue-specific content and ideological leaning of those feeds. Using these measures as our explanatory variables, we show that exposure to more topical tweets by media sources improves political knowledge, but that exposure to tweets from politicians can actually have a negative impact on knowledge. We also find weak evidence that exposure to tweets from an ideologically diverse range of sources improves political knowledge.

We proceed as follows. In Section 2, we briefly review the relevant literature before drawing on this literature to lay out our hypotheses in Section 3. We then describe both parts of our dataset – the panel survey and the tweet collection – in Section 4, and present our results in Section 5. We conclude with a discussion of implications and plans for further research in Section 6.

2 Literature

3 Hypotheses

The simplest hypothesis posits that exposure to more political information on social media should function the same way as exposure to political information in any other medium–people are inherently uninformed, and more information makes them more likely to have accurate beliefs about the world. This hypothesis does not require that all of the information on social media be true or correct, but only that it be true and correct *on average*, so that the deviations from truth become smaller in the total amount of information.

Hypothesis 1 *More exposure to political content on social media will* increase *one's level of political knowledge.*

The opposite hypothesis is also plausible: people have some accurate information about the world garnered from consuming traditional media and talking to people in their communities, but exposure to the low-quality information on Twitter makes them less informed. If Twitter is in fact an echo chamber of partisan stereotypes, biases and lies, this bad information will serve to get people farther from the truth.

Hypothesis 2 *More exposure to political content on social media will* decrease *one's level of political knowledge.*

A more nuanced theory takes advantage of the fact that we know the source of the tweets to which our respondents are exposed. It may be the case that, for die hard partisans who consciously only follow accounts that reinforce their pre-existing beliefs, political information on Twitter decreases their level of political knowledge–but also that people who follow an ideologically diverse group of Twitter accounts are better able to synthesize these conflicting viewpoints and have more correct beliefs. We operationalize this by looking at the variance of the ideology of the tweets in each person's Twitter feed.

Hypothesis 3 *Exposure to political information from an ideologically diverse group of Twitter accounts will* increase *one's level of political knowledge.*

Another aspect of the source of the tweets that we leverage is whether the account is associated with a media company/journalist or with a politician/party/campaign. We expect that the former's primary purpose is to inform, and that exposure to these kinds of tweets will increase political knowledge. On the other hand, tweets from politicians are strategically motivated, and while they may not be outright lies, they may be designed to play up fears that lead people to have less political knowledge.

Hypothesis 4 *Exposure to tweets from media accounts will* increase *one's level of political knowledge, while exposure to tweets from politicians will* decrease *one's level of political knowledge.*

4 Data

4.1 Panel Survey

We designed a 4-wave panel survey administered by YouGov to respondents drawn from a population of social media users, what YouGov calls their Social Media Analysis tool (SoMA). The SoMA sample was created by YouGov by asking respondents who had previously claimed to use social media if they would like to participate in surveys about their social media use.[1] A subset of these users who used Twitter also gave their Twitter account information to YouGov, who shared with us the Twitter timelines of each respondent. The SoMA sample contains respondents from all four nations in the United Kingdom (England, Scotland, Wales and Northern Ireland).

These respondents received a financial benefit for their participation in the survey. The surveys were conducted online using YouGov's survey module with the questions we designed for each wave, lasted around 10 minutes each, and contained between 50 and 70 questions. We supplemented these surveys with demographic information that YouGov asks of all of their respondents.

The retention rates for different waves of the survey can be seen in Table 1. Overall, there were 1,293 respondents in all 4 waves of the SoMA sample[2]. The retention was lowest between waves 1 and 2, but was otherwise similar to what

[1] The SoMA sample was maintained by YouGov to be able to link survey responses to observable happenings in on the social media world, and consists of 14,000 respondents, 7,000 each selected for their use of Twitter or Facebook. They recently changed the name of the sample to YouGov Social.

[2] In order to maintain the size of the waves, YouGov also replenished the sample, adding respondents in later waves who were not in the first wave.

is often seen in panel surveys. Notice that the retention rate is highest between waves 3 and 4 because YouGov made an intensive effort to enroll as many previous respondents for the final, post-election wave.

[Table 1 Here]

The 4 waves of the survey took place over the course of almost a year: wave 1 lasted 22 days and concluded on July 31, 2014; wave 2 lasted 8 days and concluded on December 11, 2014; wave 3 lasted 12 days and concluded on March 30, 2015; and wave 4 lasted 26 days and concluded on June 17, 2015. Wave 4 was in the field for an especially long time as part of the effort to increase the retention rate, and it began 2 weeks after the day of the general election on May 7, 2015.

The timing of the survey allowed us to measure attitudes and knowledge before, during and after the 2015 UK Parliamentary campaign and election. The "long campaign," during which spending is regulated, officially began on December 19th, 2014, and the "short campaign," in which parties are given time slots to broadcast their messages on TV, began March 30th (Hope, 2015). The Conservatives and Labour parties were the two largest parties, while the Liberal Democrats experienced a rapid decline in popularity after joining the previous coalition government with the Conservatives. The rapid rise of the UK Independence Party was a manifestation of the dissatisfaction of the nativist right with the UK's position on immigration and the EU. The election results turned out to be a huge surprise, as the Conservatives won enough seats to govern without a coalition and the Liberal Democrats were all but removed from Parliament. Despite winning 13% of the vote, UKIP won a only a single seat.

In this paper, we will be focusing on the sub-sample of SoMA who provided YouGov with their Twitter account information. While this allows us to make an inference about the impact of exposure to political information on Twitter *among people with Twitter accounts*, this is far from a representative sample of the population, and an understanding of the differences among the populations is essential. The covariate information presented in Table 2(a) was asked in each wave of the survey, and in the cases in which respondents selected different answers in different waves, the modal responses are reported.

[Table 2 Here]

Table 2(a) demonstrates that there are sizable difference between the SoMA and the voting population as a whole–the SoMA sample tends to be more male, better educated, higher socio-economic class, younger and more liberal, all of which is to be expected among social media users. The people who shared their Twitter accounts with YouGov (in the second column) are slightly more male and better educated, but in general are a reasonably representative sample of SoMA users. The data in the third column are from the British Election Study's 30,000 person post-election survey (Fieldhouse et al., 2015), and serves as the best available estimate of the true values of these demographics in the British electorate.

The SoMA respondents are also considerably more likely to prefer left-leaning parties, and more likely to have voted for Labour and especially the Green party in the 2015 election, as can be seen in Table 2(b). Our results do tend to systematically under-report support for UKIP, however. Among both samples, the breakdown by country is similar, but as shown in Table 2(c), our samples are light on respondents from Scotland and Northern Ireland and heavy on respondents from Wales.

4.2 Tweets

The "SoMA with tweets" subsection of respondents provided YouGov with their Twitter handles, and while we do not have access to their individual Twitter profiles or what they tweeted/retweeted, our novel contribution is to match the panel surveys with their Twitter timelines. The timelines consist of all of the tweets they could potentially have been exposed to during the time period from January 1st, 2014 until May 22nd, 2015[3], divided into 4 periods: from January 1st, 2014 to the beginning of wave 1 of our survey; from the end of wave 1 to the beginning of wave 2; from the end of wave 2 to the beginning of wave 3; and from the end of wave 3 until the beginning of wave 4. We thus have access to everything

[3] Excluding the days during which the surveys were actually in the field.

tweeted by every account the respondents followed.

Unlike Facebook, which uses an algorithm to tailor the order that information from friends is displayed on the user's news feed, the stream of tweets in a user's timeline is strictly chronological.[4] We cannot know which tweets among those on the timeline the user actually saw. But because the timeline is uncurated, it is reasonable to treat the tweets they saw as a random sample from all of those they might have been exposed to.[5]

To determine which tweets were politically relevant, we manually constructed short lists of terms related to our topics of interest (the British economy, Ebola, the potential Greek exit from the Eurozone, legal immigration to Britain, ties to EU, the National Health Service, and ISIS) and terms related to the four largest political parties (Conservative, Labour, Liberal Democrats, and UKIP). We then calculated which other terms frequently co-occurred with the terms in each of these lists and used these terms to expand our searches.

For example, our original search for "Ties to the EU" consisted of the terms "brexit" and "euro-skeptic"; not the most comprehensive list, but unlikely to produce many false positives. We compiled a complete dictionary of all words from all tweets, and separately, a dictionary of all words from all tweets that contained either "brexit" or "euroskeptic." We then calculated a score for each word w in this subset s:

[4]Twitter recently added a "while you were away" feature to highlight tweets that its algorithm predicts the user is likely to be interested in, but this represents a tiny fraction of the overall Twitter feed.

[5]This is actually a very tricky question unto itself, and undoubtedly there are data available that could help us do a better job of figuring out which tweets were more likely to be seen. For example, someone who only follows three people is certainly more likely to see all of their tweets than someone who follows 3,000. Similarly, holding constant the number of people being followed, someone who logs on hourly will see more tweets than someone who does monthly. Tweets during the day are probably more likely to be seen that in the middle of the night. While this remains an interesting question for future research, we think that at the individual level, taking the proportion of tweets in one's one feed on a given topic (or from a given ideological source) as a proxy for the proportion of tweets exposed to on that topic (from that ideological perspective) is reasonable as a first step.

$$Score_s^w = f_s^w f^w N_s^w$$

Where f_s^w is the relative frequency of word w in subset s, f^w is the frequency of word w overall, and N_s^w is the count of word w in subset s. We then used the words with the top 25 highest scores to create the subset of tweets that we claimed to actually pertain to the topic "Ties to the EU." The list of these terms, along with their scores can be seen in Table 3. "brexit" seems to have been an excellent choice, whereas "euroskeptic" was fairly uncommon, and more appropriate terms expressing the same sentiment included "no2eu" and "betteroffout."

[Table 3 Here]

We further refined our searches for relevant tweets into two separate categories: tweets from accounts associated with a politician or a political party (462 total accounts) and tweets from accounts associated with journalists or media outlets (987 total accounts). We calculated the number of tweets in each respondents' timeline from each of these sources about each of the topics and parties listed above.

The number of political tweets from politicians and media sources in the timelines of our respondents ranged from 0 up to 370,000. To be included in this count, a tweet needed to be (a) sent by one of the 462 political or 987 media accounts we identified and (b) mention one of the topics or parties we study. Overall, 32 percent of respondents saw 0 political tweets from either source, and 63 percent saw 0 tweets from political accounts. Indeed, tweets from media sources make up over 90 percent of the political tweets in our sample. The wide variation in this measure makes it useful as an explanatory variable. A summary of the range of the distribution is shown in Table 4.

[Table 4 Here]

The final refinement was to further divide these tweet counts by the ideology of the their source. To determine the ideology of these accounts, we used the Bayesian Ideal Point Estimation technique developed by Barberá (2015) to use the follower networks of each account to place that account on a unidimensional

left-right scale. We then coded the 25 percent leftmost journalist/media accounts as being on the left, the 25 percent rightmost journalist/media accounts as being on the right, and the remaining 50 percent as being centrist; the process was the same for the politician/party accounts.

This scoring allows us to generate an overall ideology score for the respondents' timeline by taking a weighted mean of the number of tweets in each category. This ideology score ranges from -1, which means that every tweet in the timeline came from a left-leaning source, to 1, only right-leaning sources. We can also compute the variance of the timeline–how ideologically diverse were the tweets the person might have been exposed to?

To improve the accuracy of these measures, we restrict this analysis to respondents with at least 1,000 tweets in their timelines during the period of study, leaving us with 1,598 individuals. The mean ideology scores are plotted in Figure 1(a). There are a large number of timelines with an ideology score of 0; these correspond to people who follow only centrist media and political accounts. There appear to be more people with left-leaning ideology scores, and indeed the mean ideology score (shown by the dashed horizontal line) is -.20. This value is biased downward by the large number of values at zero, so the solid horizontal line plots the mean ideology score with those 0's excluded, -.25.

[**Figure 1 Here**]

Figure 1(b) shows the diversity of the tweets in the timelines by plotting the variance of the ideology scores. The variance of all of timelines with mean ideology scores of -1, 0 or 1 have a variance of 0, hence the density of points at the left of the figure. Variance scores above .15 are uncommon, which is plausible–few people follow a diverse portfolio of political accounts, and most prefer to follow only those accounts they tend to agree with (Iyengar and Hahn, 2009). We compared these results with the responses to a question about how ideologically diverse their Twitter feeds were, and found there was essentially no correlation.[6]

[6] Our question asked them to to rank their Twitter feeds from 0 ("The accounts you follow all hold very similar views") to 100 ("The accounts you follow hold a wide variety of views"); their responses correlated with our objective measure of variance at .09. In contrast, a similar question asking them to rate the political viewpoint of their timelines (0: "Almost all the tweets you read

5 Results

5.1 Factual Knowledge

The first outcome of interest is the change in the ability of respondents to correctly answer factual questions about political topics. In waves 2 and 3 of the survey, we asked three multiple choice questions (correct answers in **bold**):

- (ISIS) The Islamic militant group known as ISIS currently controls territory in which of these countries: **Syria**, Kuwait, Morocco, or Pakistan?
- (Unemployment) Compared to a year ago, has unemployment in Great Britain increased or **decreased**?
- (Immigration) Over the past 5 years, has the number of immigrants to the United Kingdom from other EU countries been: Less than 100,000 per year, **Between 100,000 and 300,000 per year**, Between 300,000 and 500,000 per year, More than 500,000 per year?

Table 5 breaks down how many people got each question right in waves 2 and 3. ISIS was empirically the easiest question, and Immigration was most difficult; note, though, that Unemployment only had two correct answers, so people were more likely to get it correct by guessing. For both Unemployment and Immigration, we observed a greater degree of *unlearning* the correct answer than learning it– more people got the question right in wave 2 but wrong in wave 3 than vice versa. Because these questions are multiple choice, it was possible to guess the right answer, and thus some of this difference is the result of random noise. However, as discussed below, *unlearning* is consistent with H_2 and H_4, that exposure to some kinds of tweets can decrease peoples' levels of information.

[Table 5 Here]

represent a left-wing viewpoint"; 100: "Almost all the tweets you read represent a right-wing viewpoint" correlated with our measure of mean ideology at .37. Future research will explore what covariates predict accuracy in this task.)

Using the tweet counts we collected about these three topics, we tested whether people exposed to more information about one of these topics would become more likely to answer the question correctly in wave 3, controlling for their answer in wave 2. The results of our analysis can be found in Table 6. We ran three separate logistic regressions using an indicator for whether the respondent correctly answered each question as a dependent variable. We used the log of the number of tweets related to each topic in the timeline of the respondents between wave 2 and 3 of our survey, divided into those sent by Media and Politician accounts. We also included the variance of the ideology of the source of those tweets. Because the respondents are the same people in each wave of the panel, we do not include controls for any covariates.

We see that the presence of more tweets from Media sources about Unemployment and ISIS in a person's timeline between waves 2 and 3 is significantly related with an increase in knowledge about those subjects. The direction of the effect is positive and just shy of significance for Immigration. This provides support for H_1, that exposure to more political information on Twitter causes people's beliefs to become more accurate.

[Table 6 Here]

On the other hand, we see a highly significant *negative* effect of the presence of tweets from Politicians about Unemployment on respondents' knowledge. The estimated impact of tweets from Politicians in the other two regressions was weak and positive. In this particular case, this could be because the politicians tended to be using terms that we detected as being related to the (improving) unemployment rate, but which were actually more related to the much less rosy topic of austerity. While the employment rate in Britain was improving, the fierce debate being held between politicians over related economic topics like "cuts," "benefits" and "austerity" (all terms in the 25 we found related to unemployment) may have given people who saw a lot of these tweets the impression that the unemployment situation was worse than it actually was. This finding provides support for H_2, that exposure to more political information on Twitter causes people's beliefs to become less accurate. These mixed findings, supporting both H_1 and H_2, align

with our expectations outlined in H_4, as the more objective tweets sent by the media tend to improve political knowledge, while the electorally motivated tweets sent by politicians tend to harm political knowledge.

There is also a weakly significant relationship between the variance of Politician tweet ideology and improvement on the Immigration question, but the other 5 variance coefficients are insignificant and 2 of them are actually negative, so we find only weak support for H_3, that being exposed to a wider variety of ideological information increases political knowledge.

5.2 Party Placements

The other way we operationalize political information is the ability of the respondents to correctly rank the 4 major parties (Liberal Democrats, Labour, Conservatives, UKIP) on a left-right scale on three major issues in the 2015 election: Taxes/Spending, Britain's Ties to the EU, and Immigration. In each wave of the survey[7], we asked each respondent to place themselves and each of the 4 parties on a 0 (leftmost) to 100 (rightmost) scale.

One of the challenges in analysis of this sort is establishing a "ground truth" of where the parties actually stand (Tucker and Markowski, 2007). There are a wide variety of ways in which these can be done, and we measured many of them, including: the mean of all the respondents' placements of the parties; the mean of the placements by respondents with a college degree; the mean of the party placements made by self-identified supporters of each party; and the mean of the self-placements of self-identified supporters of each party.

As all of these placement estimates turned out to be highly correlated with each other at at least .93, and we decided to use the simplest measure - the mean of each placement as our "ground truth."[8] As a further reality check, we compared

[7] In wave 2 we asked these questions to half of the respondents, and in wave 3 we asked them to the other half, because of length constraints in the survey. This means that we cannot compare results from wave 2 to wave 3, but all other comparisons are possible.

[8] Among other advantages, this approach allows has the advantage of accounting for the

these placements against the party placements in the 2014 edition of the Chapel Hill Expert Survey (Bakker et al., 2015). Every wave of our placements correlated with the CHES placements at least .95. The highest correlation was with wave 1, the soonest after the 2014 survey was conducted, suggesting that differences in later waves were due to actual movements of the parties.

In order to find out the "ground truth" of where the parties should be placed, we used the aggregate responses from each wave of the survey. We used several different means to describe the true placements of each party on each in each wave:

[**Figure 2 - Figure 4**]

The histograms of placements of each party on each issue in waves 1 and 4 can be seen in Figure 2, 3 and 4. On the EU, the median ranking of the Liberal Democrats moved from 16 to 24, but the other parties stayed fairly constant. On Spending, Labour is to the left of the Liberal Democrats, and the only major movement is UKIP moving to the left. On Immigration, UKIP stayed all the way to the right, and the other 3 parties all moved to the right.

In order to determine if our respondents were "correct" in placing the parties in each wave, we used the median values of the parties as shown in Figures 2, 3 and 4. However, for the instances in which two parties were close together (within 10 points on the 100 point scale), we allowed some leeway; the correct orderings and the percentage of respondents identifying them can be seen in Table 7. Note that the correct ordering for the parties on each issue was the same in both waves for the Immigration and Spending questions, but not for the question about the EU: the Liberal Democrats moved to the right, making their position too similar to that of Labour. This meant that the EU question got "easier," hence the high percentage who got the question wrong in wave 1 but right in wave 4. Overall, the Spending question was the most difficult, with only 55 percent of respondents in wave 4 answering it correctly among those who attempted to answer it in both waves; the N is considerably smaller for this question.

movement of the parties during the campaign–notably in our case, the Liberal Democrats moved to the right on the issue of the EU.

Table 8 shows the result of the change in the ability of the respondents to accurately place the parties from Wave 1 to Wave 4. For placements on the EU issue, exposure to more media tweets increases the level of political knowledge, providing support for H_1 and H_4. However, for neither of the other issues does exposure to media tweets seem to have an effect on knowledge. Also unexpected is the negative and significant coefficient on the variance of media tweets on EU placement knowledge. Though we had hypothesized that more ideological diversity in information would allow people to better hone in on the truth, this result encourages an opposite interpretation–perhaps too many different views on the same topic can confuse people.

On the Immigration question, we find weak support for H_2, H_3 and H_4: more tweets from politicians decreases political knowledge, but higher variance in the ideology of the politicians' tweets increases political knowledge. Each of these relationships is only significant at $p < 0.10$, but notice that the sample size in these regressions is only around 500. Because of the way that these variables are constructed from the panel survey, failure to respond to one of many questions resulted in the respondent being dropped from the analysis. Future work will involve imputing missing values to take full advantage of the survey data, and these statistical relationships may be strengthened.

6 Conclusion

The problem of making inferences about what causes people to have high levels of political information is a daunting one. By using a 4-wave panel survey design and focusing on changes rather than levels, we remove the cross-sectional heterogeneity. Further, by matching survey responses to objective measures of political information on social media, we have "real-world" evidence of this process of acquiring correct information in action.

We find tentative support for the hypothesis that using social media to become politically informed causes higher levels of correct political beliefs. However, we

do find some evidence of the opposite claim: there are some sources of political information on Twitter that lead people to have incorrect political beliefs. Breaking down this analysis by the source of the tweets, our results suggest that tweets from media accounts improve political knowledge while tweets from politicians' accounts worsens political knowledge. We also find weak support for the hypothesis that following a wider variety of political accounts causes more correct beliefs.

Analysis of this kind–indeed, further analysis of the data we have collected–will add nuance to these findings. We intend to explore whether an individual's exposure to information from sources with an ideology distinct from that individual's leads to more political information. We will also look to see if the effect of political tweets from accounts that are neither media nor politicians has a similar effect.

We also have other measures of political sophistication we intend to incorporate as outcome variables–for example, before the election we asked respondents which party they thought would win their constituency. We also gave them an open-response question asking them to name the leaders of each of the major parties.

References

Bakker, Ryan, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Vachudova. 2015. "2014 Chapel Hill Expert Survey." 2015.1.

Barberá, Pablo. 2015. "Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data." *Political Analysis* 23 (1): 76–91.

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2015. "Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?" *Psychological Science* .

Carpini, Michael X Delli, and Scott Keeter. 1997. *What Americans know about politics and why it matters.* Yale University Press.

Fieldhouse, E., J. Green, G. Evans, H. Schmitt, and C. van der Eijk. 2015. "British Election Study Internet Panel Wave 6.".

Hope, Christopher. 2015. "And they're off: the 2015 general election campaign officially starts this Friday." *Telegraph UK* .

Iyengar, Shanto, and Kyu S Hahn. 2009. "Red media, blue media: Evidence of ideological selectivity in media use." *Journal of Communication* 59 (1): 19–39.

Tucker, Joshua A, and Radoslaw Markowski. 2007. Subjective vs. Objective Proximity in Poland: New Directions for the Empirical Study of Political Representation. 2007 Annual Meeting of the American Political Science Association.

Table 1: Retention Rates Among Survey Respondents

Sample	Wave 1	Wave 2	Wave 3	Wave 4	All Waves
NR respondents	1,118	1,047	1,094	958	1,660
Retention, previous wave		63%	71%	87%	465 (in all 4 waves)
SoMA respondents	2,574	2,507	2,776	2,490	3,846
Retention, previous wave		68%	79%	90%	1,308 (in all 4 waves)

Retention rates were high, and there were 1,308 respondents in the SoMA sample that completed all 4 waves of the survey. Note that wave 4 is the only post-election wave.

Table 2: Descriptive Statistics of Relevant Populations

Panel A: Covariates

	SOMA	SOMA w tweets	BES
Women	45%	43%	50%
15+ Years Education	52%	55%	41%
Median Age	48	48	53
Median HH Income	£34,200	£37,500	£27,500
Median L-R Ideology†	5.2	5.2	4.6

† Self-reported ideology, left to right; asked on a 0-100 scale in our survey and on a 0-10 scale in the BES. The BES is a nationall representative post-election survey of 30,000 voters1.

Panel B: Vote Choice, Post-Election

	SOMA	SOMA w tweets	Election
Conservative	33	32	37
Labour	34	35	31
Liberal Democrats	8	9	8
SNP	5	5	5
UKIP	9	8	13
Green	10	11	4
Other	1	1	3
	100%	100%	100%

Panel C: UK Country

	SOMA	SOMA w tweets	Reality
England	84	85	84
Scotland	5	5	8
Wales	9	9	5
Northern Ireland	1	1	3

The demographic, vote choice and geographic vote share of the relevant populations: the Social Media Analaysis sample, and the subgroup for whom we have their Twitter timeline.

Table 3: Top 10 Terms Pertaining to the Topic "Ties to the EU"

Term	Score
brexit	1000
no2eu	44
betteroffout	18
eureferendum	6.7
eu	6.7
euref	5.9
grexit	2.2
scoxit	1.5
stayineu	1.3
flexcit	1.3
...	...
euroskeptic	.25

Examples of the terms we found to tend to co-occur with our anchor terms for the topic "Ties to the EU." We used this process to find terms that identify a tweet as pertaining to a topic of interest.

Table 4: Summary of Counts of Political Tweets in Respondents' Timelines

	1st Quartile	Median	Mean	3rd Quartile	Max
Politician tweets	0	0	2,774	1,784	166,500
Media tweets	0	5,066	22,350	28,260	307,600
All Relevant tweets	0	5,978	25,130	30,630	370,800

The unit of analysis here is the respondent, and the summary statistics are of the number of tweets from politician and media accounts that appeared in their timeline. Sample are the 2481 SoMA respondents

Table 5: Distribution of Responses to Knowledge Questions

	ISIS		Unemployment		Immigration	
	Right W2	Wrong W2	Right W2	Wrong W2	Right W2	Wrong W2
Right W3	88%	5%	51%	10%	31%	17%
Wrong W3	3%	4%	13%	26%	20%	32%
Total W2	91%	9%	64%	36%	51%	49%

Cell entries are percentages for each possible combination of right and wrong answers across wave 2 and wave 3 of the knowledge questions: (R,R), (R,W), (W,R), (W,W). Bottom line shows how difficult each question was showing the percentage correct in wave 2. Sample is the 1,226 respondents who provided access to Twitter timelines, and who had a chance to respond to the indicated questions in wave 2 and wave 3.

Table 6: Effects of Exposure to Topical Tweets on Changes in Knowledge

	Unemployment	ISIS	Immigration
Log tweets by Media on Topic	0.106**	0.227**	0.0709
	(0.0457)	(0.104)	(0.0455)
Log tweets by Politician on Topic	-0.146***	0.171	0.0485
	(0.0537)	(0.413)	(0.0640)
Variance of Tweet Ideology by Media	0.331	-0.842	-0.423
	(0.624)	(1.249)	(0.540)
Variance of Tweet Ideology by Politician	1.009	4.930	1.448*
	(0.853)	(5.744)	(0.812)
Question Correct Previous Wave	2.305***	3.120***	1.101***
	(0.139)	(0.264)	(0.120)
Constant	-1.023***	0.0222	-0.772***
	(0.125)	(0.216)	(0.101)
N	1225	1225	1225

Standard errors in parentheses. The dependent variable in each logit regression is an indicator for whether the respondent got that question correct in wave 3. Analysis of tweets includes only the those tweeted between wave 2 and wave 3 of our survey.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Placement of Parties in Waves 1 and 4

	EU, N= 1,220	
Correct Order W1	LibDem < Labour < Conservatives < UKIP	
Correct Order W4	LibDem = Labour < Conservatives < UKIP	
	Right W1	Wrong W1
Right W4	52%	26%
Wrong W4	5%	17%
	Immigration, N= 1,197	
Correct Order W1	Labour = LibDem < Conservatives < UKIP	
Correct Order W4	Labour = LibDem < Conservatives < UKIP	
	Right W1	Wrong W1
Right W4	62%	14%
Wrong W4	10%	14%
	Spending, N= 937	
Correct Order W1	Labour < LibDem < Conservatives = UKIP	
Correct Order W4	Labour < LibDem < Conservatives = UKIP	
	Right W1	Wrong W1
Right W4	36%	19%
Wrong W4	16%	29%

Cell entries are the percentage who got each placement question correct in wave 1 and wave 4, among the respondents who gave us their Twitter timelines. The correct ordering for the parties on each issue was the same in both waves for the Immigration and Spending questions, but not for the question about the EU: the Liberal Democrats moved to the right, making their position too similar to that of Labour.

Table 8: Effects of Seeing Topical Tweets on Changes in Party Placement

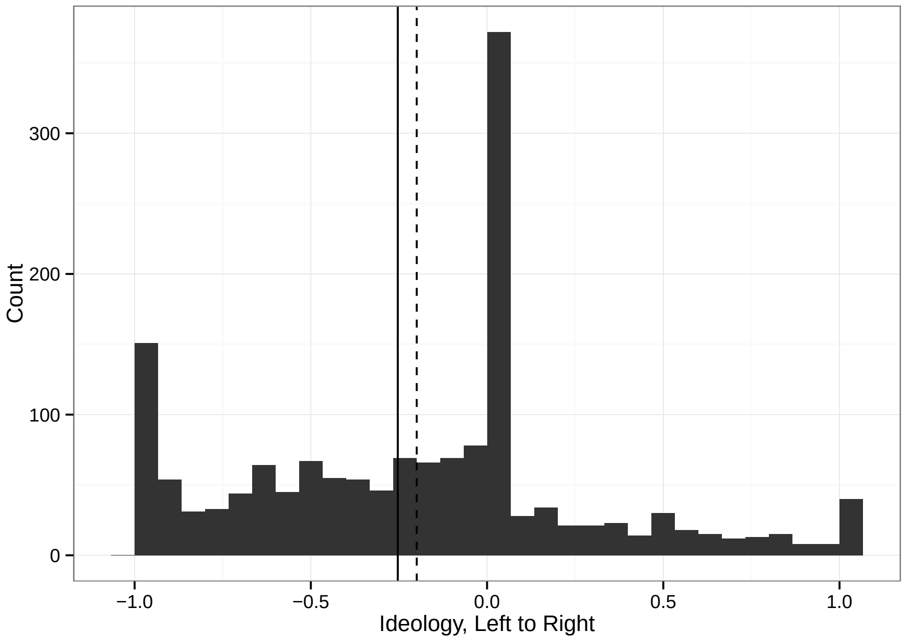
	EU	Econ	Immigration
Log Politician Tweets	-0.0162	0.0332	-0.119*
	(0.0933)	(0.0424)	(0.0507)
Log Media Tweets	0.355**	-0.0459	0.0470
	(0.116)	(0.0496)	(0.0534)
Variance of Tweet Ideology by Media	-2.640**	1.036	0.854
	(0.862)	(0.768)	(0.861)
Variance of Tweet Ideology by Politician	0.916	0.496	6.065*
	(1.356)	(0.803)	(2.545)
Placement Correct Wave 1	2.160***	1.188***	1.846***
	(0.289)	(0.214)	(0.257)
N	515	408	505

Standard errors in parentheses. The Dependent Variable in each logit regression is an indicator for whether the respondent got that question correct in Wave 1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

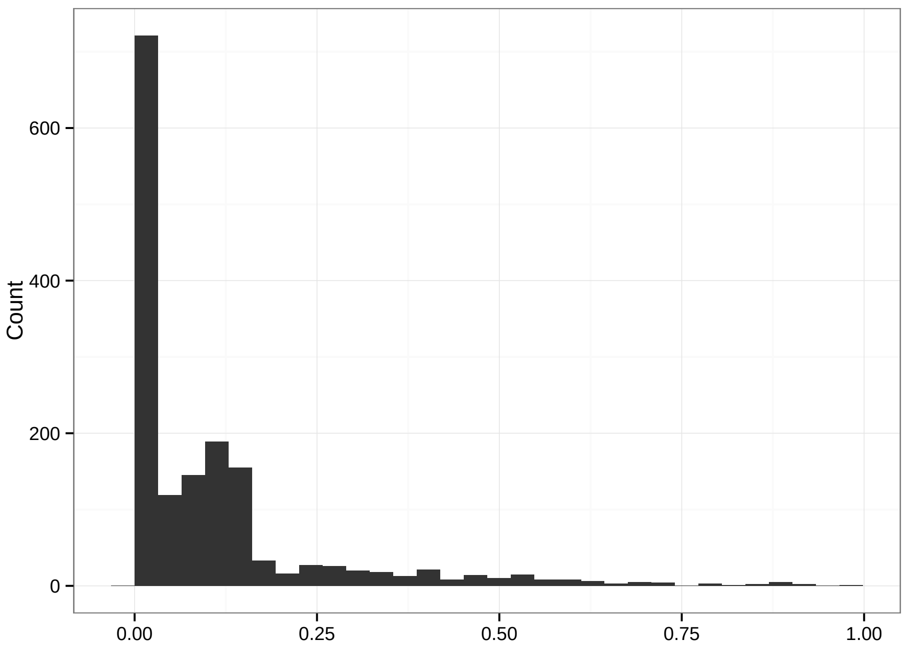
Figure 1: Distribution of Tweet Ideology, N=1,598

A: Mean Ideology Scores



The mean ideology of the political tweets sent by Politician and Media accounts in the timelines of those respondents with at least 1,000 such tweets. The dashed horizontal line is the average ideology score, and the solid horizontal line is the average ideology score excluding the observations with an average ideology score equal 0.

B: Variance of Ideology Scores



The variance of the ideology of those political tweets.

Figure 2: Party Placement, Wave 1 to Wave 4: EU

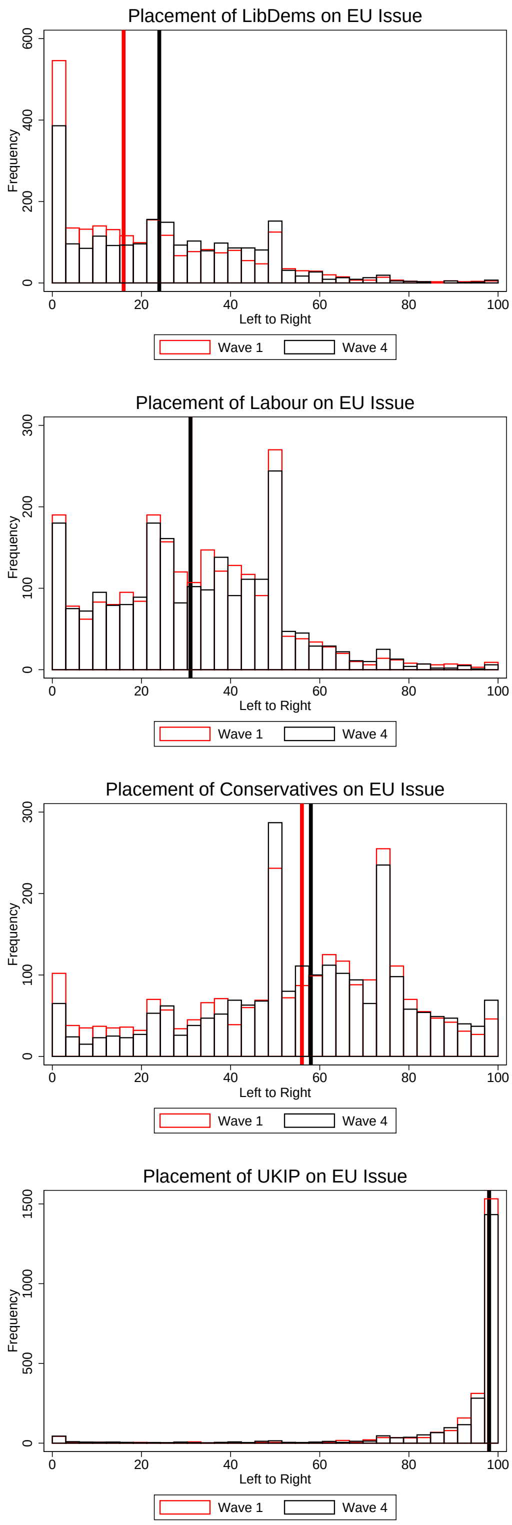


Figure 3: Party Placement, Wave 1 to Wave 4: Spending

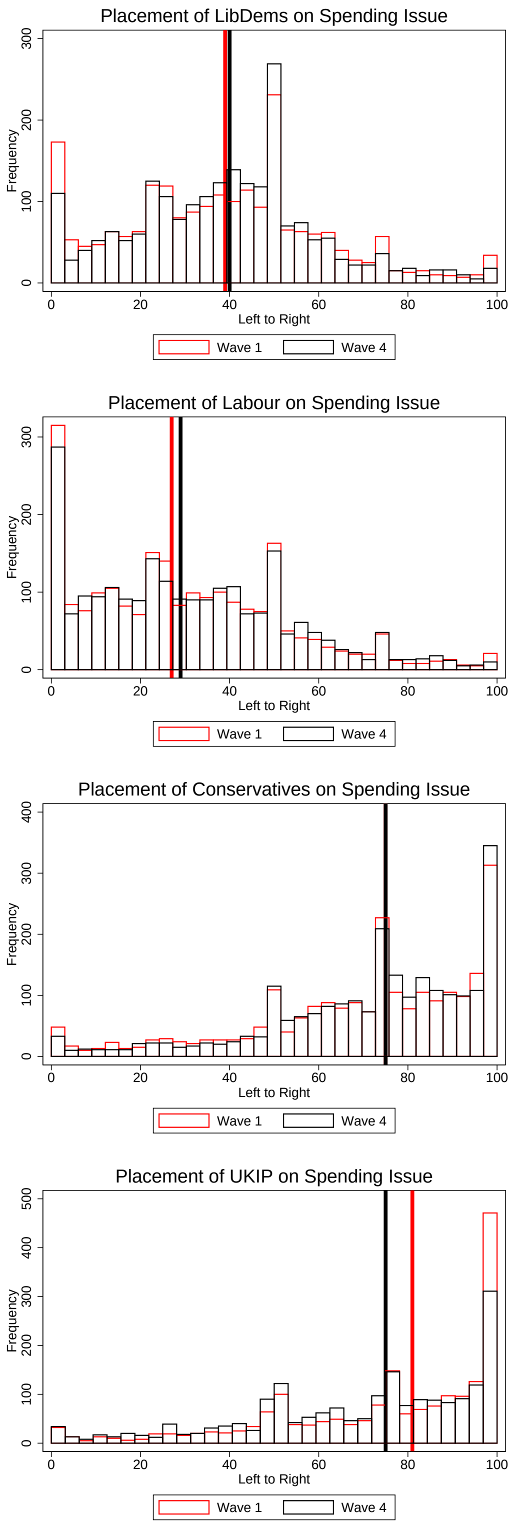


Figure 4: Party Placement, Wave 1 to Wave 4: Immigration