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

Kevin Munger *

December 15, 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 the effect of exposure to political tweets on changes in subjects' political knowledge. We show that tweets sent by media accounts tend to increase knowledge of factual questions, and that tweets sent by politicians tend to increase knowledge of party platforms. We also examine issues specifically relevant to the 2015 UK election, and find that tweets by incumbent parties improve estimates of the state of the economy while tweets by opposition parties diminish them, and that tweets by UKIP about immigration tend to inflate beliefs about the number of immigrants to the UK.

*With Patrick Egan, Jonathan Nagler, Jonathan Ronen, and Joshua A. Tucker. 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–or under what circumstances. 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 (Bakshy, Messing, and Adamic, 2015; Barberá et al., 2015)?

In this paper, we test a series of hypotheses related to the ways in which exposure to political information on social media affects political knowledge. We expect the effect of this exposure to be generally positive, and for the effects to be larger for information sent by media outlets and political parties on factual knowledge and party platforms, respectively. 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 (Delli Carpini and Keeter, 1997). Furthermore, people looking for political information on social media might also be looking for political information elsewhere.

To address these concerns, we synthesize two types of data in our research design. 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 – changes 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. Although the accounts that each user chooses to follow are the product of self-selection and thus somewhat the result of ideological homophily, those accounts tweet about topics that are at least partially exogenous to this selection process, and by comparing changes in the levels of political knowledge of the same subject across different topics, we obtain a causal estimate of the impact of exposure to political information about a specific topic.

We find evidence that exposure to tweets from accounts associated with UK parties leads to an increase in the ability to correctly rank the parties' platforms on relevant issues, but is not related to an increase in factual knowledge. Important exceptions are highly salient, politicized issues: tweets from UKIP (an anti-immigration party) tend to increase estimates about the rate of immigration, tweets from incumbent parties tend to decrease estimates of the rate of unemployment, and tweets from opposition parties tend to increase estimates of the rate of unemployment. On the other hand, tweets from media accounts are generally associated with increased factual knowledge, but not with the ability to correctly rank the parties' platforms.

## 2 Literature

There is a general consensus that a more politically informed citizenry causes a better-functioning democracy (Campbell et al., 1960; Converse, 1964). On an individual level, political knowledge is a key ingredient in aligning preferences with the political behaviors most likely to realize those preferences. Although there is some scholarly debate about whether and to what extent cognitive shortcuts enable voters who are less knowledgeable to still make the correct voting decision (Bartels, 1996; Fowler and Margolis, 2014; Lau and Redlawsk, 2001; Lupia, 1994), there is little doubt that increased political knowledge contributes to the individual and collective functioning of democracy.

And levels of political knowledge have been improving. The US population has become better informed in the past several decades, especially in terms of "surveillance knowledge"–facts about issues of current political interest (Delli Carpini and Keeter, 1991, 1997). Though the topic is less well-studied, there is evidence of a similar increase in knowledge of topical political issues in the UK (Tillman, 2012).

The biggest driver of changes in the flows of political information in society as a whole is the media environment. The development and expansion of new media technologies, from newspapers to broadcast TV to cable, have changed both the amount of political information in the media environment and the motivations of media corporations. Different media are easier or more difficult for voters to learn from; Jerit, Barabas, and Bolsen (2006) find that newspaper coverage of a topic tends to increase the gap in knowledge between the more and less educated about that topic, while Curran et al. (2009) find that "public service" television broadcast systems like in Denmark tend to produce a more informed citizenry relative to the more market-based system in the US and UK.

The mere exposure to political information in the media, however, is far from sufficient to produce more knowledgeable citizens. The "perceptual screen" of Campbell et al. (1960) has been developed into a theory of motivated cognition: people process ideologically consonant information faster than dissonant information (Redlawsk, 2002); people are overly skeptical of ideologically dissonant information and insufficiently skeptical of consonant information (Taber and Lodge, 2006); and people are more likely to have higher levels of information about topics that are consistent with their worldviews (Jerit and Barabas, 2012). There is even evidence of a "backlash effect" of being corrected about false but ideologically comfortable beliefs–these corrections cause a greater partisan affiliation and confidence in the false beliefs (Nyhan and Reifler, 2010; Nyhan et al., 2014).

The effects of partisan motivated cognition are especially pronounced when people self-select into consuming ideologically agreeable media, a phenomenon for which there is strong evidence (Iyengar and Hahn, 2009; Stroud, 2008). Even though people are aware of the biased nature of the news they consume, they are still persuaded by it, as has been widely studied in the case of cable news networks Fox News and MSNBC in the US (DellaVigna and Kaplan, 2006; Martin and Yurukoglu, 2014). With the advent of social media, the ability of people to further personalize their information environments has caused concern that political communication is increasingly taking place in an "echo chamber." While early studies supported this worry (Conover et al., 2012), more recent work that does not rely on self-selected samples has shown that a large amount of cross-partisan political exchange does take place (Bakshy, Messing, and Adamic, 2015; Barberá et al., 2015).

# 3 Hypotheses

Because there are various types of political information about which we ask our subjects, we have different hypotheses about how they will respond to political tweets.

One type of political information we measure is explicitly partisan. We ask each respondent to place each of the major parties on a left-right spectrum on three separate issues: the EU, spending/taxation, and immigration. By communicating about these topics, the parties are stating their own position, often in the context of the position of the other parties. We expect that being exposed to tweets sent by the party's official accounts or politicians from that party on a specific topic will be related to an increased ability to estimate the parties' stance on that issue.

**Hypothesis 1** *Exposure to information sent by a political party on a specific topic will be related to an increase in knowledge of the parties' relative stances on that issue.*

On the other hand, tweets about these political topics sent by the media are likely to be concerned more with facts about that issue, rather than about the parties' stances.

**Hypothesis 2** *Exposure to information sent by a media organization on a specific topic will* ***not*** *be related to an increase in knowledge of the parties' relative stances on that issue.*

The other type of political information in our study concerns factual, multiple-choice questions about political topics. Because the media is discussing the facts concerning these topics, we expect to see a relationship between exposure to more tweets sent by the media and an increase in this factual knowledge.

**Hypothesis 3** *Exposure to information sent by a media organization on a specific topic will be related to an increase in knowledge of the facts associated with that issue.*

Tweets sent by politicians are strategically motivated, and while they may not be outright lying, they are certainly more concerned with projecting competence and conveying their stances on issues than with communicating factual information. As a result, our expectation for them is the converse of above.

**Hypothesis 4** *Exposure to information sent by a political party on a specific topic will* ***not*** *be related to an increase in knowledge of the facts associated with that issue.*

There are a few notable exceptions to this hypothesis. Certain facts are inherently political, in that they reflect the competence of the incumbent parties or the gravity of other highly salient issues. The primary metric by which voters evaluate incumbent parties is through economic performance. During the tenure of the incumbent parties, unemployment had been steadily falling, a fact that the current government (the Conservatives and the Liberal Democrats) wants to be widely known and the opposition (UKIP and Labour) wants to obscure. We expect that exposure to tweets about unemployment sent by incumbent parties to be associated with an increase in knowledge about the true level of unemployment, and for the opposite to be true for opposition parties.

Another important issue in contemporary British politics is legal immigration from the EU to the UK. Concerns about the rate of immigration were instrumental in the rise of the UK Independence Party (UKIP), a party whose anti-immigration stance resonated with a large number of voters. We expect that UKIP will want to draw as much attention to the issue as possible, and that exposure to tweets about immigration sent by UKIP will increase knowledge about the rate of immigration.

**Hypothesis 5** *Exposure to information sent by certain political parties on strategically advantageous topics* ***will*** *be related to an increase in knowledge of the facts associated with those issue.*

# 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 re-

[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.

spondents 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 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. Also, wave 4 consists only of respondents who had participated in at least one of the previous three waves, to best take advantage of the panel design.

**[Table 1 Here]**

The four 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

[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.

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 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 or retweeted, our novel contribution is to match the panel surveys with their Twitter timelines. The timelines consist of all of the tweets to which they could potentially have been exposed 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 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] We can, however, incorporate self-reported information on the subject's Twitter use to improve these measures. For example,

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

[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.

we would expect someone who claims to use Twitter "About once a day" to be more likely to be exposed to the political tweets in her timeline than someone who uses Twitter "Once or twice a week." In the regression analyses performed below, we weight respondents based on their answer to a six-point ordinal ranking of their Twitter use. We also always control for the total number of tweets in the subjects' timelines, so that the explanatory variable of interest is actually the *percentage* of political tweets.

To determine which tweets were politically relevant, we manually constructed short lists of terms related to our topics of interest (British economic policy, Ebola, unemployment in the UK, legal immigration to Britain, ties to EU, the National Health Service, and ISIS) and terms related to the four largest political parties (Conservatives, 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 $s$:

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

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 performed an additional categorization of relevant tweets based on the type of the account that created them: 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 further broke the political accounts into those associated with each of the four major political parties under study. For media accounts, we identified the media organizations with the most high-profile Twitter accounts in the UK and divided them according to their ideological leanings. Major left-leaning media outlets are The Guardian and The Independent; right-leaning media outlets are The Financial Times and The Sun; centrist media outlets are the STV, the BBC, CNN and The Times.

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. The wide variation in this measure makes it useful as an explanatory variable. A summary of the distribution of the tweets in the subjects' timelines is shown in Table 4.

**[Table 4 Here]**

Each cell of Panel A shows the number of subjects who saw at least one tweet sent by a type of account about each topic of interest. Comparing rows of Panel A shows the relative "penetration" of each party/media type among our subjects: we see that Labour and the Conservatives, the two largest parties, have tweets that reach the most subjects, and that centrist media reaches the most subjects overall.

Panel B restricts its summary to only those subjects who show up in Panel A, for seeing at least one relevant tweet. Each cell of Panel B shows the mean number of tweets sent by each party/media type about each topic in the subjects' timelines, after removing all of the subjects for which this value is zero. This analysis does not attempt to address the number of separate accounts in each tweeting bloc; for example, it might be the case that each person who follows at least one Left Media account is likely to follow exactly one Left Media account, but that each person who follows at least one Labour account is likely to follow multiple Labour accounts; this makes the inference from comparing the rows of Panel B less obvious.

The important comparison in Panel B is between the columns. Based on a largely fixed set of Twitter accounts, we see the relative importance each tweeting bloc places on each topic. Relevant cells are in bold: UKIP emphasizes Immigration and the EU, while Labour is concerned with the Economy. The Left Media avoids the topic of Immigration, and the Centrist Media loves talking about ISIS.

Essentially, Panel A describes the breadth of the 28 (7 blocs $\times$ 4 topics) tweetments[6] under study, and Panel B describes the dosages. This distinction is important for our identification strategy: the variation in Panel A is due to self-selection, but that in Panel B is less so. The topics that the accounts that our subjects have chosen to follow can vary independently of this choice, and because all of our measurements of tweet counts are topic-specific, at least some of the variation in exposure to tweets is exogenous.

# 5 Results

## 5.1 Party Placements

The first outcome of interest 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)

[6] Sorry.

[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, and in practice, we find that there is too little power to use the results from waves 2 and 3 in our analysis.

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 potential measures of this ground truth, and we tested 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 .93 or higher, 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 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 estimates 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.

Plotting histograms of the density of party placements allows us to both compare between parties in the same wave and track the movement of the parties from wave 1 to wave 4. These histograms of each party on each issue can be seen in Figures 1, 2 and 3.

[**Figure 1 - Figure 3**]

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

[8]Among other advantages, this approach allows for tracking the movement of the parties during the campaign. Notably, the Liberal Democrats moved to the right on the issue of the EU, and all of the parties except UKIP moved to the right on immigration

in each wave, we used the median values of the parties as shown in Figures 1, 2 and 3. 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.

The relevant results of the statistical tests we run are presented in Figure 4. Each of the twelve horizontal lines is the coefficient (with standard errors) of the logged number of tweets[9] in the subject's timeline sent by an account affiliated with that party *and* related to the that topic between waves 1 and 4. The dependent variable in each case is whether the subject correctly ranked the four parties on that topic in wave 4 of the survey; because this is binary, it is estimated with a probit model. In order to estimate the *change* in knowledge, we control for whether they correctly ranked the four parties on that topic in wave 4. Each regression includes a number of other demographic control variables,[10] as well as controls for the number of tweets in the subject's timeline sent by accounts affiliated with one of the major media sources mentioned above and the total number of tweets in the subject's timeline. These regressions also weight the respondents by their

[9]Throughout the analysis, we use (1.0001 plus) the log of the number of tweets in subject's timelines because of the highly skewed nature of these distributions; for brevity's sake, we refrain from saying "log of" in the rest of the paper.

[10]Standard demographic controls are gender, age, class (using the British 5-category system), years of education, race (a binary variable for "white British" or not), marital status, religiosity (binary). Specific control variables for other patterns of media consumption we add are frequency of watching long-running news program Newsnight and frequency of using the internet, (both on a 5-point, ordinal scale), and dummy variables for Newspaper Type. In the UK, different types of newspapers are significant signifiers of group identity and carry different kinds of news, so reading "Red Tops" (tabloids like the The Sun or The Daily Mirror) or "Broadsheets" (The Guardian or The Telegraph) is an important measure of media exposure.

self-reported frequency of Twitter use.

We see in Figure 4 that seven of the twelve effects are positive and significant at $p < .05$, and one more at $p < .10$, providing support for $H_1$. In general, the effects on ranking the parties on their stance on the EU are poorly estimated due the relatively few tweets about the EU in our sample.

By far the largest effect is that of tweets by UKIP about immigration, in line with UKIP's well-known attention to the topic. As noted above, UKIP was located at the far right of the ideological spectrum on the issue of immigration for the duration of the survey, and the other three parties all moved noticeably to the right between wave 1 and wave 4. These facts are consistent with UKIP's criticism of the mainstream parties for being soft on immigration.

In order to better compare the effects of tweets by the different parties–and to display the effects of media tweets–we conduct analyses that combine these measures into the same regression. The results in Figure 5 are derived from the same model specifications as above, except that each of the three regressions contains the tweet count from each of the four parties instead of just one. The graphs in this figure also show the effect of tweets sent from media outlets.

In general, we find moderate support for both $H_1$ and $H_2$. Only two of the nine possible estimates of the effect of media tweets are significant, and they are in opposite directions. On the other hand, six of the possible twelve effects of party tweets are positive and significant at $p < .05$, and two more at $p < .10$; none are negative and significant. Additionally, the weakest effects are on the topic of the EU. As seen in Figure 4 Panel B, the EU was the least tweeted-about of these three topics, so these smaller effect sizes are plausible.

## 5.2 Factual Knowledge

The other way we operationalize political information 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?[11]
- (Unemployment) Compared to a year ago, has unemployment in Great Britain increased, **decreased**, or stayed the same?
- (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 3 rather than 4 possible 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.

**[Table 5 Here]**

The regressions we run to explore the determinants of changes in this form of political knowledge use the same specification as above. We ran three separate probit regressions using an indicator for whether the respondent correctly answered each question as a dependent variable.[12] The results can be found in Figure 6. The explanatory variables are again the number of tweets sent by each of listed sources on the relevant topic, with the same demographic, media use, and previous level of knowledge controls as above.

[11] In the Wave 2 version of this question, "Morocco" was "Egypt," and we made the switch because there some news reports of ISIS activity in Egypt after Wave 2.

[12] We coded "Don't Know" responses as incorrect.

**[Figure 6 Here]**

We find weak support for $H_3$: four of the nine estimated effects of media tweets are positive and significant, and one is negative and significant. We find more support for $H_4$, as only two of the twelve estimated effects are positive and significant.[13] In fact, four of the twelve are *negative* and significant.

We also find support for $H_5$, about the highly salient and politicized topics. As expected, tweets about Unemployment by the incumbent Tories increase knowledge that unemployment had in fact decreased, while tweets from Labour, the main opposition party, decreases that knowledge. The status of the Liberal Democrats is somewhat confusing, because although they were technically part of the incumbent government, they suffered a significant loss of support because of their decision to cooperate with the more conservative Tories on economic policy. Regardless, the negative and significant effect of their tweets was unexpected.

We were also surprised to see no effect of UKIP tweets on knowledge of immigration. We attempt to explain this result: because the categorical responses to the questions about Immigration and Unemployment are ordered, we can analyze the change in the *absolute level* of the subject's estimates of the rate of unemployment and immigration, rather than the correctness of that estimate.

**[Table 6 Here]**

Table 6 displays the results of ordered probit regressions using these ordered categorical responses as the dependent variable, so that higher levels of the DV are associated with an increase in the estimate of the rate of unemployment or immigration.[14] We use the same suite of demographic and media use controls as above, and weight for frequency of Twitter use.

[13] The effect of UKIP on ISIS is poorly estimated because of near-perfect separation: no one who saw tweets from UKIP about ISIS got the question wrong.

[14] For example, the dependent variable takes a value of 1 if the subject switched their response from "Unemployment has decreased" to "Unemployment has stayed the same" in the first column and a value of 2 if she switched from "Between 100,000 and 300,000 [immigrants] per year" to "More than 500,000 [immigrants] per year" in the second column.

The results of column 1 concord with the evidence from Figure 6: tweets from Labour increase estimates of the change in the rate of unemployment, and tweets from Tories decrease those estimates. Because the truth is that unemployment had been decreasing, this corresponds with less and greater accuracy, respectively. And while column 2 does not provide support for $H_5$, it displays a logically consistent result: tweets from UKIP increase changes in the estimates of the rate of immigration. If some of these changes were from a value that was too low to the correct value and others were from the correct value to a higher value, the net effect on accuracy would be insignificant, as in Figure 6.

One way to corroborate this story is to look at the heterogeneous effects on different subgroups based on their ideology. Figure 7 displays the results of the same analysis on the Unemployment factual question as in Figure 6, except that it divides the sample into liberals and conservatives based on their self-reported ideology scores. The median value of this variable[15] is 43.5 out of 100, which is where we split the sample. This value is to the left (where the left is associated more with being more liberal) of the median of the range of the question, but since the population we study (Twitter users) skews to the left, 43.5 is the appropriate cut point.

[**Figure 7 Here**]

There are contrasting results in Figure 7. Generally, there are fewer significant results, due to the reduced sample size. But the results that remain significant are that tweets from UKIP increase accuracy among liberals, while tweets from Labour increase accuracy among conservatives. This is precisely what we would expect to see: tweets from parties to which the subjects are ideologically opposed tend to make them more knowledgeable about this politicized factual question.

Overall, we find moderate support for our hypotheses, and try to explain the cases in which support is lacking. Exposure to political information sent by parties tends to increase knowledge of party platforms but not of factual knowledge, while the converse is true for information sent by media accounts. Political par-

[15]The question is asked in waves 2, 3 and 4 of the survey, and the value we use is the average over all the waves in which the subject responded to the question.

ties do, however, tend to affect factual knowledge of heavily politicized issues in strategically coherent ways.

# 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.

Our findings largely bear out our initial expectations, although they do admit several concerns. First, our findings deal only with a non-representative sample of social media users, and we have not yet addressed the issue of how our results would generalize, either to the entire population of social media users or to the UK as a whole. Second, we still cannot be sure which of the tweets in the subjects' timelines they actually saw; we attempt to address this by controlling for their self-reported frequency of Twitter use, but this measure is far from perfect. Finally, our identification strategy is not airtight: we observe a correlation between exposure to tweets and changes in knowledge about related topics, but we cannot be sure that tweets are the true cause. Future work, with a newly-expanded dataset of subjects, could pool the changes over different periods and include person fixed effects, to better take advantage of the panel design and get at this causal question.

# 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.

Bakshy, Eytan, Solomon Messing, and Lada Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* p. aaa1160.

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* .

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

Campbell, Angus, Philip E Converse, Warren E Miller, and E Donald. 1960. "The American Voter.".

Conover, Michael D, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2012. "Partisan asymmetries in online political activity." *EPJ Data Science* 1 (1): 1–19.

Converse, Phillip. 1964. "The Nature of Belief Systems in Mass Publics. In Ideology and Discontent, ed. David Apter. New York: Free Press.".

Curran, James, Shanto Iyengar, Anker Brink Lund, and Inka Salovaara-Moring. 2009. "Media System, Public Knowledge and Democracy A Comparative Study." *European Journal of Communication* 24 (1): 5–26.

DellaVigna, Stefano, and Ethan Kaplan. 2006. The Fox News effect: Media bias and voting. Technical report National Bureau of Economic Research.

Delli Carpini, Michael X, and Scott Keeter. 1991. "Stability and Change in the US Public's Knowledge of Politics." *Public Opinion Quarterly* 55 (4): 583–612.

Delli Carpini, Michael X, 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.".

Fowler, Anthony, and Michele Margolis. 2014. "The political consequences of uninformed voters." *Electoral Studies* 34: 100–110.

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.

Jerit, Jennifer, and Jason Barabas. 2012. "Partisan perceptual bias and the information environment." *The Journal of Politics* 74 (03): 672–684.

Jerit, Jennifer, Jason Barabas, and Toby Bolsen. 2006. "Citizens, knowledge, and the information environment." *American Journal of Political Science* 50 (2): 266–282.

Lau, Richard R, and David P Redlawsk. 2001. "Advantages and disadvantages of cognitive heuristics in political decision making." *American Journal of Political Science* pp. 951–971.

Lupia, Arthur. 1994. "Shortcuts versus encyclopedias: information and voting behavior in California insurance reform elections." *American Political Science Review* 88 (01): 63–76.

Martin, Gregory J, and Ali Yurukoglu. 2014. Bias in cable news: Real effects and polarization. Technical report National Bureau of Economic Research.

Nyhan, Brendan, and Jason Reifler. 2010. "When corrections fail: The persistence of political misperceptions." *Political Behavior* 32 (2): 303–330.

Nyhan, Brendan, Jason Reifler, Sean Richey, and Gary L Freed. 2014. "Effective messages in vaccine promotion: a randomized trial." *Pediatrics* 133 (4): e835–e842.

Redlawsk, David P. 2002. “Hot cognition or cool consideration? Testing the effects of motivated reasoning on political decision making.” *The Journal of Politics* 64 (04): 1021–1044.

Stroud, Natalie Jomini. 2008. “Media use and political predispositions: Revisiting the concept of selective exposure.” *Political Behavior* 30 (3): 341–366.

Taber, Charles S, and Milton Lodge. 2006. “Motivated skepticism in the evaluation of political beliefs.” *American Journal of Political Science* 50 (3): 755–769.

Tillman, Erik R. 2012. “Support for the euro, political knowledge, and voting behavior in the 2001 and 2005 UK general elections.” *European Union Politics* 13 (3): 367–389.

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.

Panel A: Number of Subjects With At Least One Apropos Tweet

| | ISIS | EU | Economy | Immigration |
|---|---|---|---|---|
| Labour | 568 | 579 | 579 | 579 |
| Tory | 469 | 506 | 506 | 504 |
| LibDem | 220 | 245 | 246 | 246 |
| UKIP | 110 | 110 | 110 | 110 |
| Right Media | 149 | 155 | 155 | 155 |
| Left Media | 159 | 163 | 162 | 162 |
| Centrist Media | 714 | 780 | 732 | 774 |

Panel B: Mean Apropos Tweets in Timeline Among Those With At Least One

| | ISIS | EU | Economy | Immigration |
|---|---|---|---|---|
| Labour | 131 | 775 | **2,561** | 1,781 |
| Tory | 35 | 325 | 747 | 470 |
| LibDem | 26 | 350 | 715 | 470 |
| UKIP | 48 | **2,142** | 1,003 | **2,561** |
| Right Media | 27 | 197 | 512 | 421 |
| Left Media | 24 | 114 | 132 | **93** |
| Centrist Media | **122** | 366 | 540 | 529 |

Table 4: Panel A summarizes how many subjects were exposed to tweets by the entities and about the topics under study. For example, the top left numerical cell explains that at least one tweet about ISIS sent by an account affiliated with Labour appeared in the timeline of 588 subjects. There are 2,755 subjects for whom we have timeline access. Panel B displays the mean number of tweets in the timelines of those subjects summarized in Panel A. For example, the bottom right corner says that, among the 774 subjects with at least one tweet about Immigration sent by Centrist Media in their timeline, the mean number of such tweets in their timelines is 529. Cells bolded for attention.

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). The 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 responded to the indicated questions in wave 2 and wave 3.

**Effect of Tweets on Changes in Absolute Levels of Unemployment/Immigration**

| | *Dependent variable:* | |
|---|---|---|
| | Δ Unemployment | Δ Immigration |
| Labour Tweets | $0.076^{**}$ | $-0.030^{\dagger}$ |
| | (0.020) | (0.018) |
| UKIP Tweets | 0.022 | $0.071^{*}$ |
| | (0.044) | (0.031) |
| LibDem Tweets | 0.021 | −0.014 |
| | (0.037) | (0.036) |
| Tory Tweets | $-0.058^{\dagger}$ | 0.038 |
| | (0.033) | (0.029) |
| Right Media Tweets | −0.027 | −0.020 |
| | (0.042) | (0.059) |
| Right Media Tweets | 0.001 | 0.008 |
| | (0.033) | (0.031) |
| Right Media Tweets | $-0.077^{*}$ | −0.022 |
| | (0.031) | (0.032) |
| Demographic controls | ✓ | ✓ |
| Media Use controls | ✓ | ✓ |
| Observations | 1,111 | 1,111 |
| *Note:* | | $^{\dagger}$p<0.1; $^{*}$p<0.05; $^{**}$p<0.01 |

Table 6: Estimates of the impact of the number of tweets in the subject's timeline sent by an account affiliated with that party or group of media outlets *and* related to the that topic, calculated from twelve separate regressions. The dependent variable in each case is an ordered factor variable that corresponds to the answer the subject gave to that factual question in wave 3, estimated with an ordered probit model. There are three categories for the Unemployment question (arranged in increasing order of the estimate of the change in unemployment) and four categories in the Immigration category (arranged in increasing order of the estimate of the number of immigrants); "Don't Know" is coded to the median category. Each regression includes demographic and media consumption control variables, as well as a control for the response of the subject in wave 2. Each regression weighted the respondents by their self-reported frequency of Twitter use.

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.

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

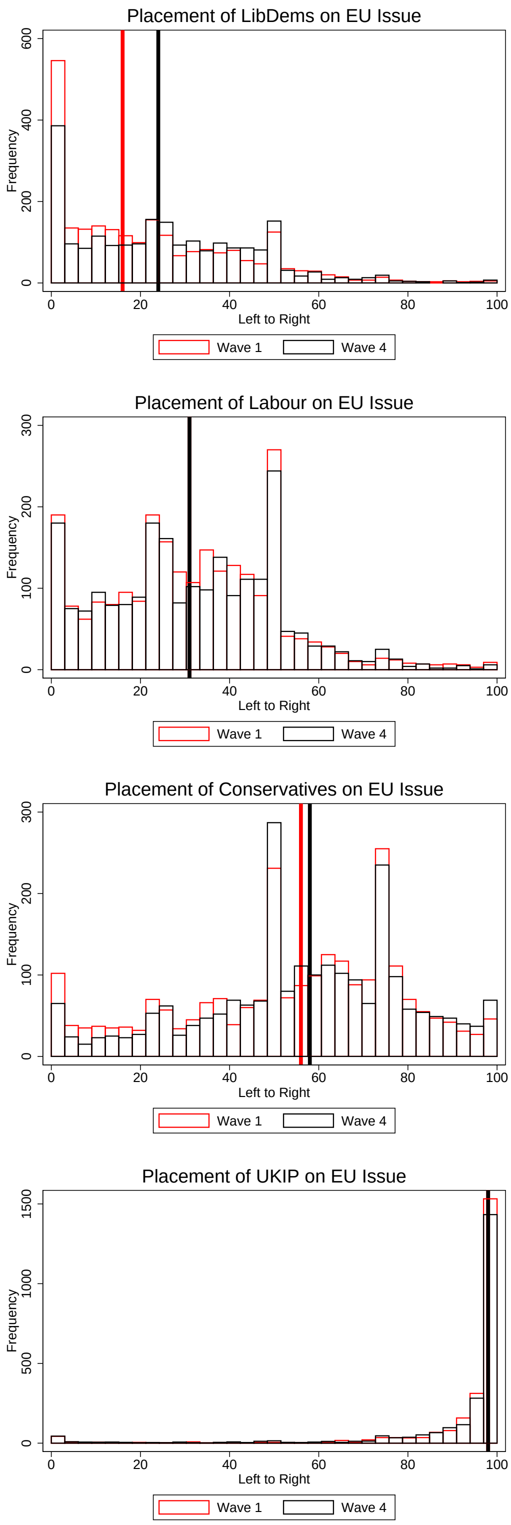


Figure 2: Party Placement, Wave 1 to Wave 4: Spending

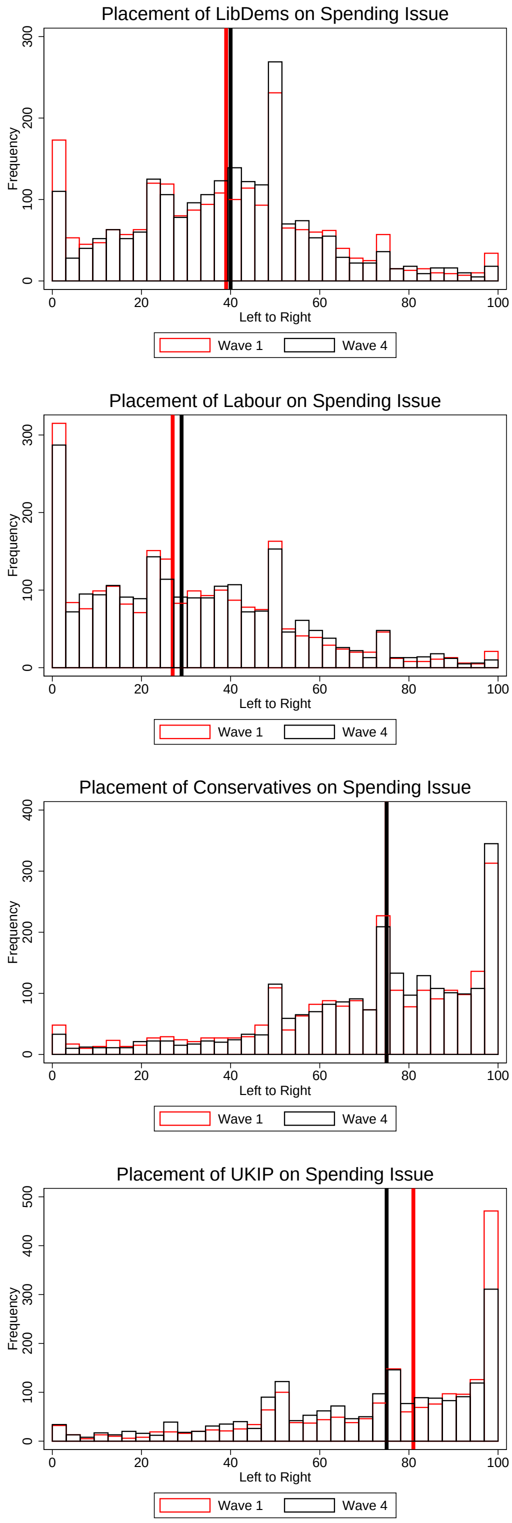
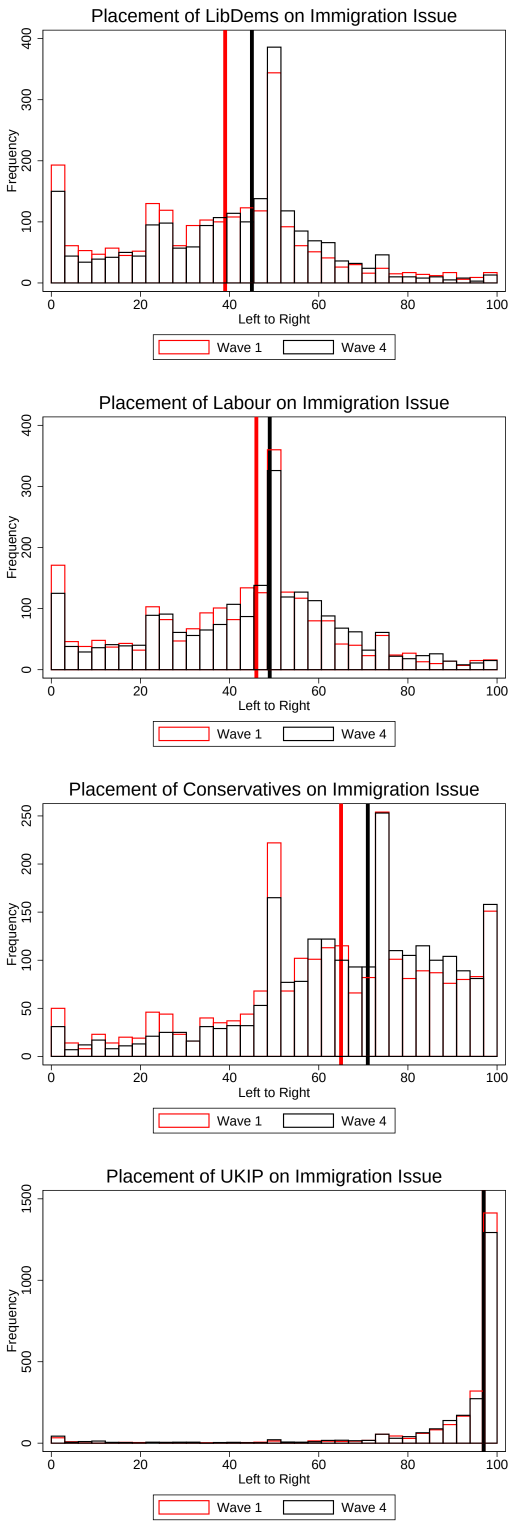


Figure 3: Party Placement, Wave 1 to Wave 4: Immigration



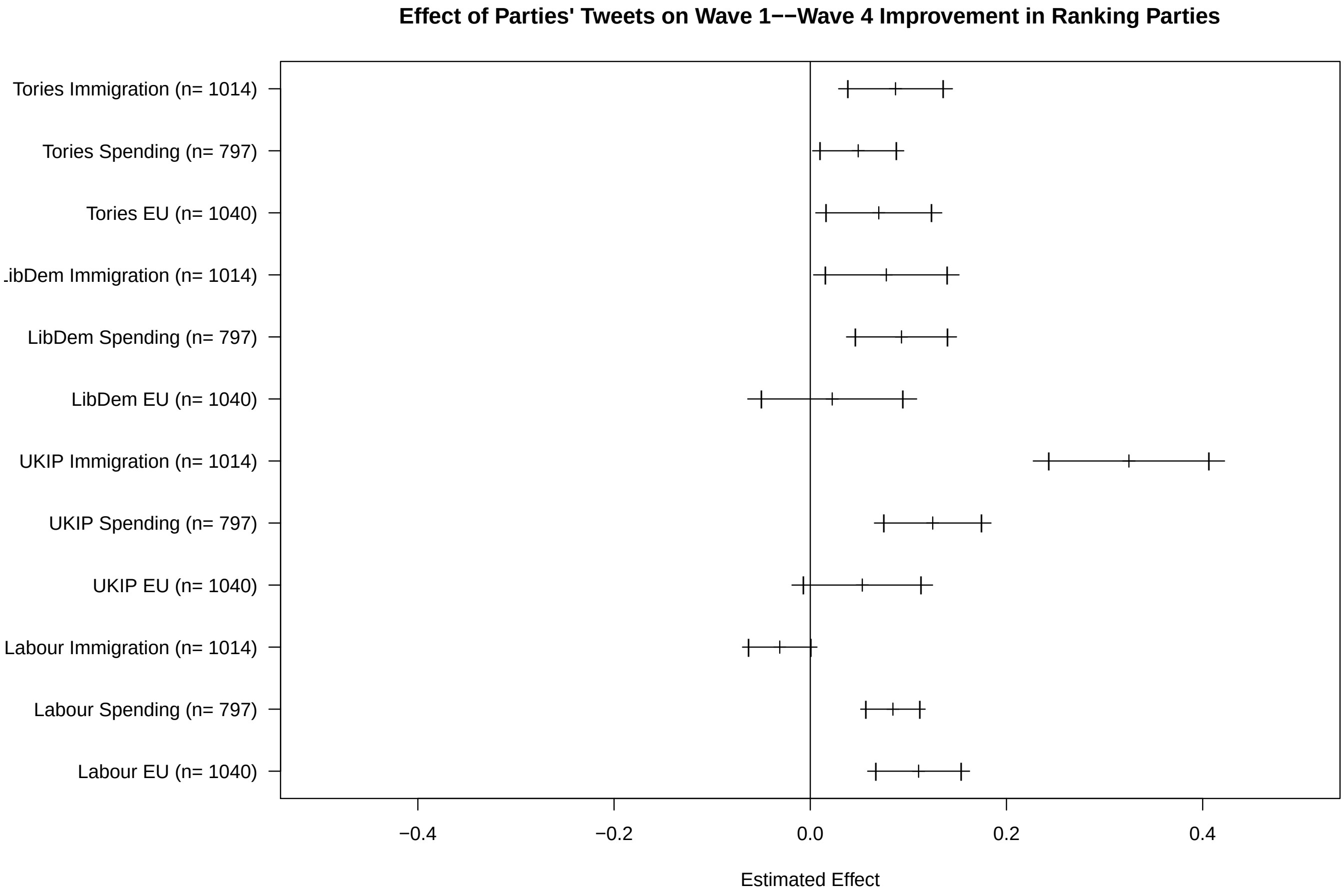
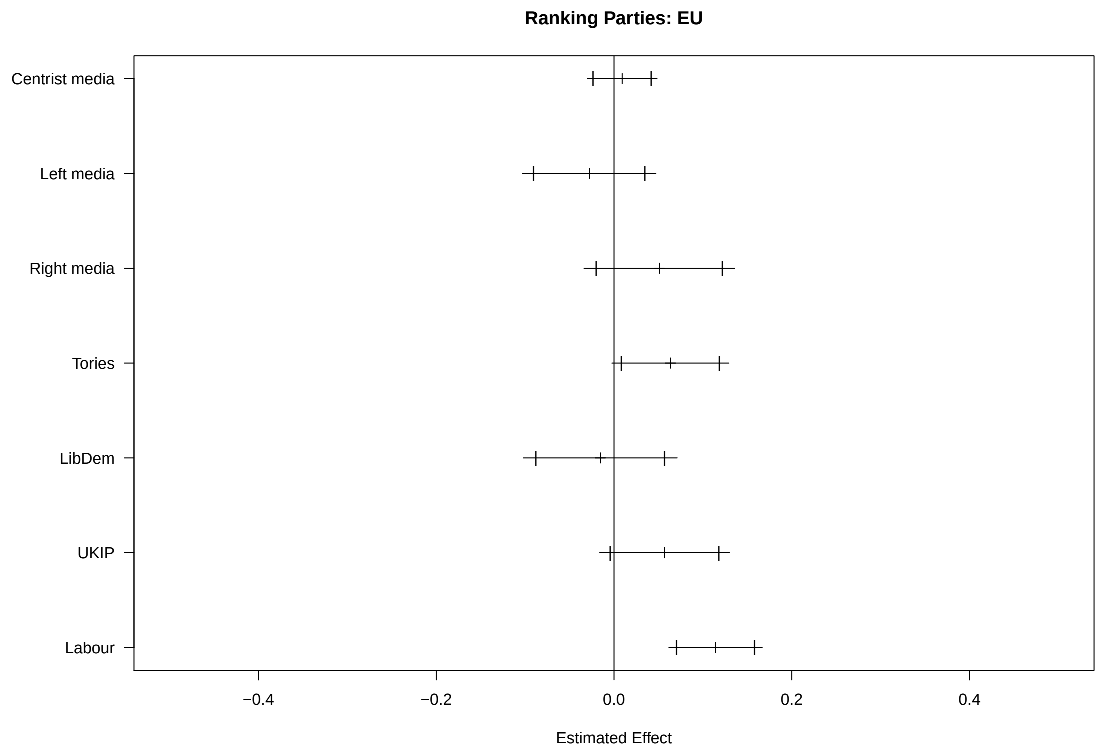
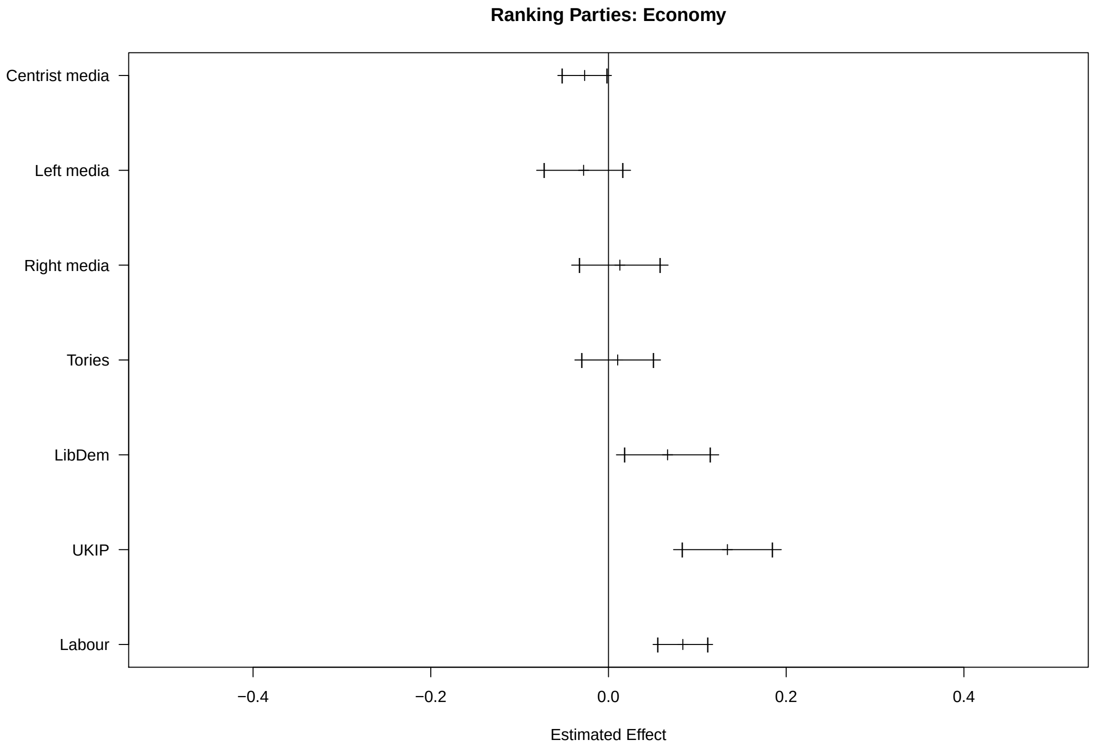



Figure 4: Estimates of the impact of the number of tweets in the subject's timeline sent by an account affiliated with that party *and* related to the that topic. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the subject correctly ranked the four parties on that topic in wave 4 of the survey; because this is binary, it is estimated with a probit model. Each regression includes demographic, media consumption, and media tweets control variables, as well as a control for whether the subject correctly ranked the parties on that topic in wave 1. Each regression weights the respondents by their self-reported frequency of Twitter use.

Figure 5: Estimates of the impact of the number of tweets in the subject's timeline sent by an account affiliated with that party or media outlet *and* related to the that topic. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the subject correctly ranked the four parties on that topic in wave 4 of the survey; because this is binary, it is estimated with a probit model. Each regression includes demographic, media consumption, and media tweets control variables, as well as a control for whether the subject correctly ranked the parties on that topic in wave 1. Each regression weights the respondents by their self-reported frequency of Twitter use.




Ranking Parties: Economy
Centrist media
Left media
Right media
Tories
LibDem
UKIP
Labour
−0.4
−0.2
0.0
0.2
0.4
Estimated Effect

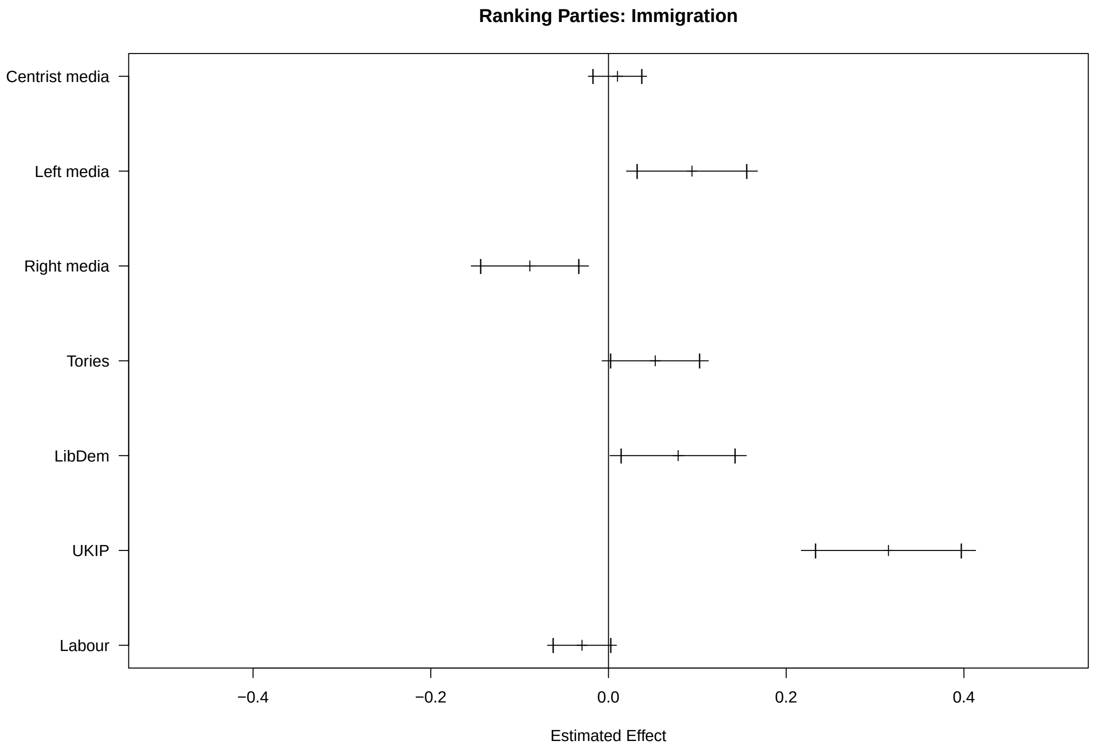
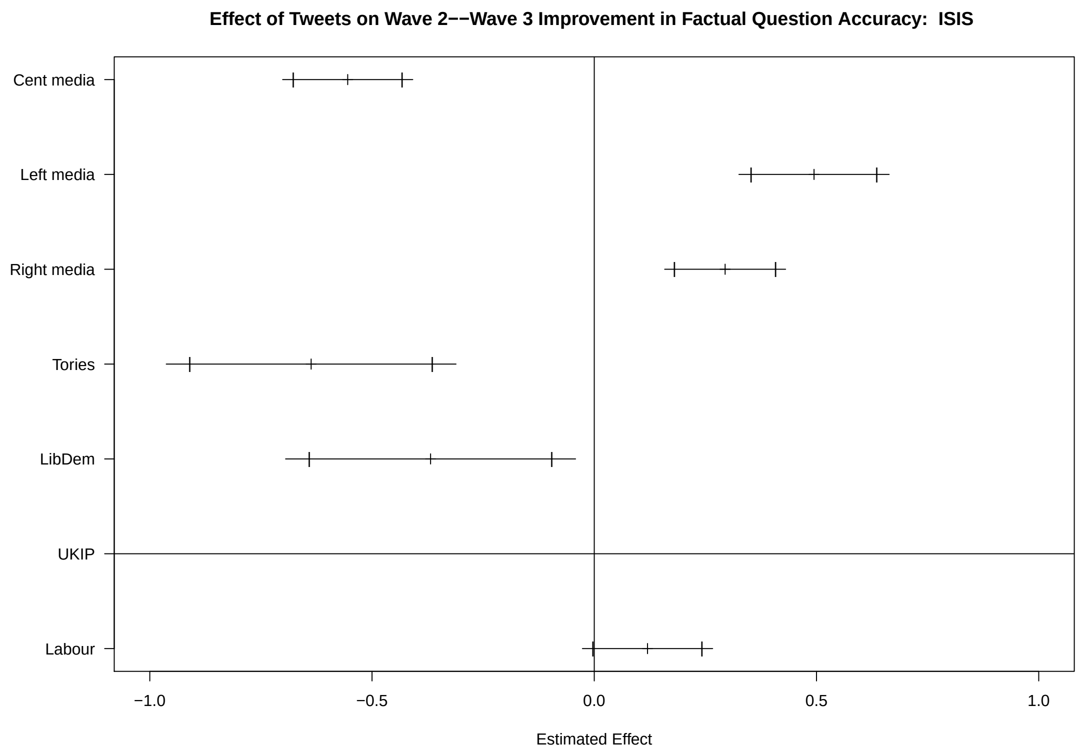
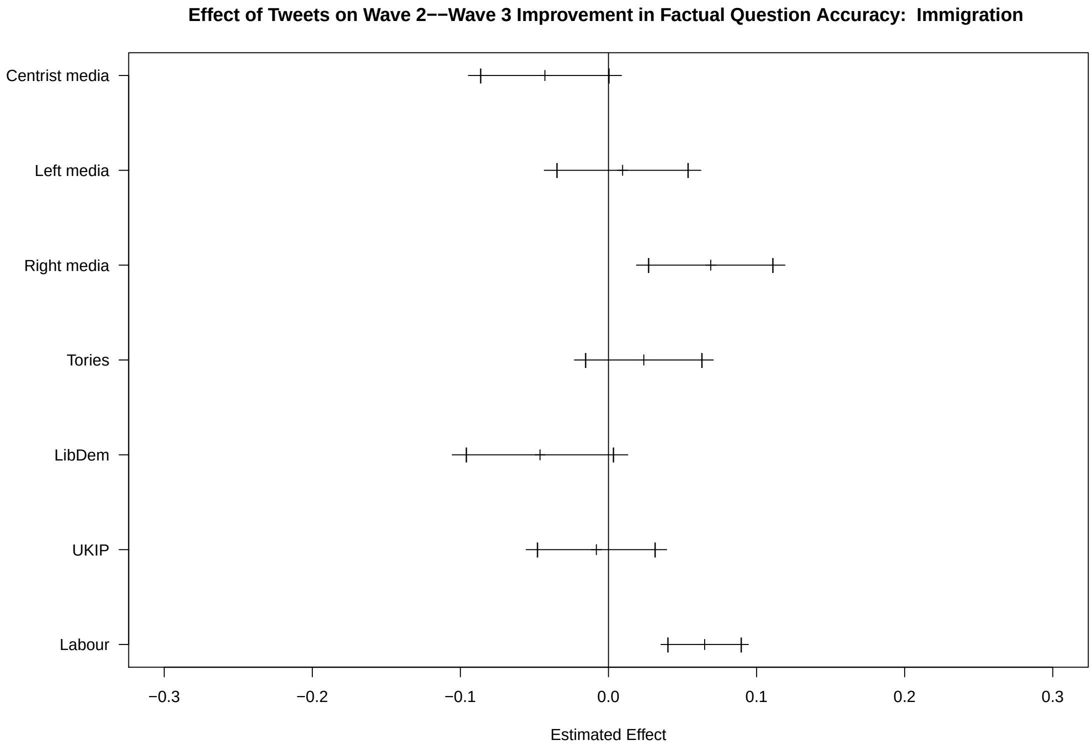

Ranking Parties: Immigration
Centrist media
Left media
Right media
Tories
LibDem
UKIP
Labour
−0.4
−0.2
0.0
0.2
0.4
Estimated Effect

Figure 6: Estimates of the impact of the number of tweets in the subject's timeline sent by an account affiliated with that party or media outlet *and* related to the that topic between waves 2 and 3. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the subject correctly answered the factual multiple choice question on that topic in wave 4 of the survey; because this is binary, it is estimated with a probit model. Each regression includes demographic, media consumption, and media tweets control variables, as well as a control for whether the subject correctly answered factual question in wave 2. Each regression weights the respondents by their self-reported frequency of Twitter use. Notice that the x-axis for the ISIS graph is more zoomed-out than for the other two.




Effect of Tweets on Wave 2−−Wave 3 Improvement in Factual Question Accuracy:  Immigration
Centrist media
Left media
Right media
Tories
LibDem
UKIP
Labour
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
Estimated Effect

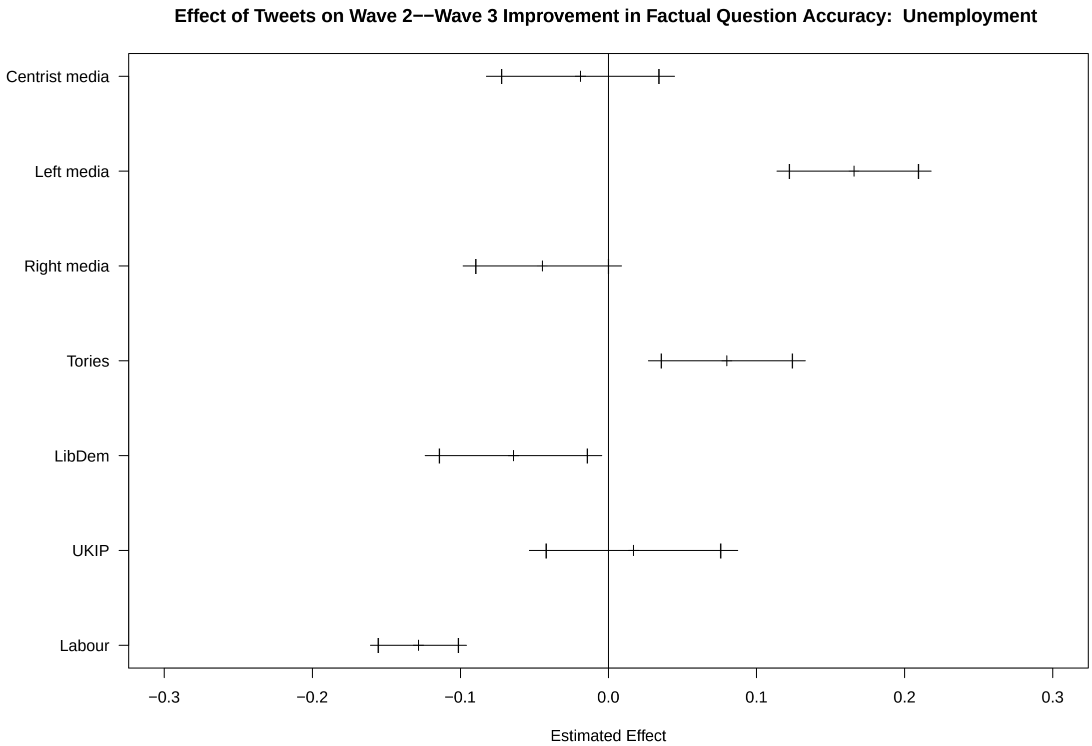

Effect of Tweets on Wave 2−−Wave 3 Improvement in Factual Question Accuracy:  Unemployment
Centrist media
Left media
Right media
Tories
LibDem
UKIP
Labour
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
Estimated Effect

Figure 7: Restricted to conservatives and liberals, respectively, estimates of the impact of the number of tweets in the subject's timeline sent by an account affiliated with that party or media outlet *and* related to the that topic between waves 2 and 3. Subjects are divided into liberals and conservatives by splitting the sample at the median value of self-reported ideology, which was 43.5 out of 100, with liberals associated with lower values. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the subject correctly answered the factual multiple choice question on that topic in wave 4 of the survey; because this is binary, it is estimated with a probit model. Each regression includes demographic, media consumption, and media tweets control variables, as well as a control for whether the subject correctly answered factual question in wave 2. Each regression weights the respondents by their self-reported frequency of Twitter use.