Learning (and Unlearning) from the Media and Political Parties: Evidence from the 2015 UK Election[*]

Kevin Munger[†] Patrick Egan[‡] Jonathan Nagler[§]
Jonathan Ronen,[¶] and Joshua A. Tucker[‖]

August 4, 2016

DRAFT IN PROGRESS; PLEASE CONTACT AUTHORS FOR MOST UP TO DATE VERSION BEFORE CITING

Abstract

Political knowledge is crucial to the functioning of democratic political systems, and therefore represents an important topic of study for political scientists. Attempts to measure determinants of political knowledge, however, have long relied on self-reports of key causal variables (such as media exposure) and often must contend with concerns regarding endogeneity. To address both these issues simultaneously, we conduct a panel survey that spans the 2015 UK election to measure changes in political knowledge and rely on a *non-self-reported* measure of media exposure. More specifically, we analyze respondents Twitter feeds to objectively measure exposure to political information, including the source of that information. We show that information from media accounts tends to increase knowledge of factual questions, and that information from politicians tends to increase knowledge of party platforms. We also examine issues specifically relevant to the 2015 UK election, and find that information from incumbent parties improves estimates of the state of the economy while tweets by opposition parties diminish them, and that information from UKIP about immigration tends to inflate beliefs about the number of immigrants to the UK.

1 Introduction

For democratic political systems to function, citizens must have at least a minimal amount of information about the political process. Many determinants of political knowledge are intrisic to the individual–both gender and education are strongly predictive of an individual's level of knowledge (Barabas et al., 2014). However, the media and technological environment plays a role in aggregate levels of knowledge, with heterogeneous effects on individuals based on their media exposure (Prior, 2007). Compared to gender and education, though, exposure to media is difficult to measure, and self-reports are unreliable (Prior, 2012).

Media technology is rapidly changing, and for an increasing number of citizens, social media is an important source of exposure to political news. Politicians and parties tweet their views directly to citizens, while media outlets and journalists have another tool at their disposal to reach the public outside the confines of articles and broadcast. However, we are only in the earliest days of beginning to understand the causal effect of social media usage on political participation, especially in established democracies. In particular, we still know very little about whether or not social media usage causes people to become more informed about politics, and, if so, 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 could affect political knowledge. We expect the effect of exposure to be generally positive, but also that these effects will vary by the source of the information. In particular, we predict that information from political parties will have a stronger effect on voters ability to correctly place parties' positions on issues relative to one another, while we expect information from the media to increase factual knowledge about politically relevant events.

Testing causal theories of the effect of social media use is essentially impossible using only 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, and gender – and simply observe the way that people become more (or less) politically knowledgeable during the course of the campaign.[1] 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. There is strong evidence that self-reports of exposure to news on traditional media are flawed (Prior, 2009), and recent findings show that this is true of self-reports of Twitter use as well (Scharkow, 2016). 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 their 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 respondent across different topics, we obtain a causal estimate of the impact of exposure to political information about a specific topic.

---

[1] By examining changes rather than levels of information, we avoid the problem identified in Prior and Lupia (2008) that knowledge levels measured through unmotivated surveys can underestimate true levels of knowledge.

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 Partisan Information and Learning

There is a general consensus that a more politically informed citizenry is associated with 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 short-cuts 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.

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 (Carpini and Keeter, 1991; Delli Carpini and Keeter, 1997). Though the topic is less well-studied, there is evidence of a similar increase in knowledge of topical political issues, and of increases in knowledge of party platforms during campaigns, in the UK (Andersen, Tilley and Heath, 2005; Tillman, 2012). Partisan politics in the UK have been undergoing a large shift in the new millennium. Poor economic performance and the unpopular war in Iraq in the mid to late 2000s led to the

decline of New Labour, the term for the move of the Labour Party to the right on economic policy headed by Prime Minister Tony Blair. (Whiteley et al., 2013). The incumbent government in 2015 was a coalition between the Conservatives and the much smaller Liberal Democrats. The coalition government turned out poorly for the Liberal Democrats, who lost support among the left for their cooperation with the Conservatives and especially for their support (contrary to their campaign promises) of an increase in university tuition fees (Weaver, 2015). Much of this support in Scotland and Wales, which had been Liberal Democrat strongholds, transferred to national parties: the Scottish National Party, and Plaid Cymru in Wales. Distance between the traditional parties and voters on the issues of immigration and the EU similarly hurt the Liberal Democrats and Labour and helped give rise to the nativist UK Independence Party (UKIP) (Evans and Mellon, 2015). This makes the 2015 UK Parliamentary election an excellent case with which to study political learning, as politics is highly salient and information about the parties in flux.

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, changes both the amount of political information available and distribution of content in a given citizen's media bundle (Prior, 2007). These changes affect people in different ways. 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 Prior (2005) find that cable and the Internet increase knowledge among consumers of news media but not among those who prefer entertainment. There is also evidence from cross-national comparisons of a "general equilibrium" effect of media environments, with "public service" television broadcast systems in Denmark and Switzerland tending to produce a more informed citizenry relative to the more market-based system in the US and UK (Curran et al., 2009; Iyengar et al., 2009).

The mere exposure to political information in the media, however, is far from sufficient to produce more knowledgeable citizens. This idea of a partisan "per-

ceptual screen" that goes all the way back to Campbell et al. (1960), but has been given a more robust theoretical treatment as the theory of motivated cognition has been applied to politics: 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" suggesting that when false but ideologically comfortable beliefs are corrected, these corrections actually leader to an even 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). Using an approach that is similar to the one we employ in this paper, Flaxman, Goel and Rao (2013) measure the ideological diversity of news and opinion articles in Twitter timelines. They find some ideological segregation, but that this effect is smaller on Twitter than for articles accessed via search engines.

3 Hypotheses

Based on the theories of media effects and motivated cognition discussed above, we have distinct hypotheses about how respondents will learn from different types of political tweets.[2]

First, we test the hypothesis that exposure to more tweets about a particular issue will be associated with an increase in knowledge about that issue. Although we believe that characteristics of the source and respondent will mediate this learning, we expect the aggregate effect to be positive.

Hypothesis 1 *Exposure to information on Twitter about a political topic will be positively related to an increase in knowledge about that topic.*

"Political knowledge" has been measured with a variety of approaches, ranging from questions about long-standing institutions ("How many members in the House of Commons?") to questions about current policy proposals ("Does Cameron's recently proposed immigration bill aim to increase or decrease the number of immigrants to Britain?"). In a recent article in the *American Political Science Review*, Barabas et al. (2014) show that exposure to media can have heterogeneous effects on these different measures of political knowledge. Specifically, they find positive effects of media exposure only on "surveillance" knowledge (as opposed to static civic knowledge). Further, these effects are restricted only to general surveillance knowledge, as opposed to policy surveillance knowledge.

We operationalize "knowledge" in two ways. The first is explicitly partisan, and constitutes policy surveillance knowledge. We ask each respondent to place each of the four major parties on a left-right spectrum on three issues: the UK's ties

---

[2]Note that our design is incapable of perfectly identifying the causal effect of exposure to tweets on political knowledge, and we avoid using causal language in this section. Our approach will inform even the most epistemologically conservative researcher about the way that different types of citizens learn during the course of the campaign. We believe that there is an important direct effect of political information in tweets on knowledge, and while we control for some measures of traditional media consumption, we cannot differentiate the effect of political tweets from all of the other unmeasured sources of information to which the respondents are exposed.

to the EU, spending/taxation, and immigration to the UK. As will be explained in greater detail in the following section, a "correct response" is one that *correctly order the party from left-right* on that issue. We assume it is in a given party's interest to attempt to differentiate themselves from other parties on these issue dimensions, and thus the cumulative effect of exposure to Tweets from parties should be a better understanding of whether the parties stand *vis a vis* each other on these issues.[3] This expectation concords with the findings in Andersen, Tilley and Heath (2005), that knowledge of party platforms increased during election campaigns in the UK. Thus our second hypothesis is:

Hypothesis 2 *Exposure to information on Twitter about a political topic sent by a political party (but not a media organization) will be related to an increase in knowledge of the parties' relative positions on that issue.*

The other type of information in our study is measured through factual, multiple-choice questions about political topics, examples of which are provided below. These are the general surveillance knowledge questions for which Barabas et al. (2014) find positive media effects. Here we expect that parties might very well have an incentive to distort information about facts to better serve their partisan interests, and that the aggregate effect of exposure to information from the parties will be negligible:

Hypothesis 3 *Exposure to information on Twitter sent by a media organization (but not a political party) on a specific topic will be positively related to an increase in knowledge of the facts associated with that issue.*

While we do not generally expect tweets from political parties to drive *correct* knowledge of political facts, we can still generate some more topic-specific hypothe-

---

[3]To be clear, this means we are *not* assuming a Downsian world whereby these is some "correct" position on each issue for maximizing votes. If that were the case, then we would expect the exact opposite effect: additional information from parties should make it harder for voters to correctly order the parties, as each party would locate themself on this ideal spot and scatter the remaining parties far away. Readers who prefer this set of assumptions could consider this opposite prediction a null hypothesis.

ses. Certain facts are inherently political, in that they reflect the competence of the incumbent parties or the gravity of other highly salient issues. For example, it has long been posited that voters evaluate incumbent parties to a large extent through economic performance (Lewis-Beck and Paldam, 2000). As unemployment had been steadily falling in the years leading up to the 2015 UK election, we assume that the incumbent parties – the Conservatives and the Liberal Democrats – would want this fact to be widely known and the opposition parties (Labour and UKIP) would want to obscure this fact. Thus we get predictions in different directions for different parties on this issue: 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 UKIP, a party whose anti-immigration stance resonated with a large number of voters. UKIP wanted to draw as much attention to the issue as possible, and we expect that exposure to tweets about immigration sent by UKIP would increase knowledge about the rate of immigration. An equally plausible hypothesis would be that exposure to UKIP tweets would simply *increase* one's estimate of the total number of immigrants, as opposed to increase the likelihood of one knowing the *correct* number of immigrants.

Hypothesis 4 *Exposure to information sent by certain political parties on strategically advantageous topics* **will** *be positively related to an increase in knowledge of the facts associated with those issue: (A) Tweets from incumbent parties will increase knowledge about changes in unemployment; (B) Tweets from opposition parties will decrease knowledge about changes in unemployment; (C) Tweets from UKIP will increase knowledge about the correct rate of immigration to the UK; (D) Tweets from UKIP will increase belief in the number of immigrants coming to the UK.*

4 Data

4.1 Panel Survey

We designed a 4-wave panel survey administered by the polling firm 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.[4] 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. To preserve anonymity, YouGov did not share the actual Twitter accounts of the respondents. The SoMA sample contains respondents from all four countries 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 retained for all 4 waves of the SoMA sample, out of the 3,846 who appeared in at least wave [5]. The retention was lowest between waves 1 and 2, but was otherwise similar to what is often seen in online panel surveys (Chang and Krosnick, 2009). 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

---

[4]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.

[5]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.

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 experienced a sharp decline in popularity after joining the previous coalition government with the Conservatives. The rise of UKIP was a manifestation of the dissatisfaction of the nativist right with the UK's position on immigration and the EU (Evans and Mellon, 2015). The election results turned out to be a surprise, as pre-election polls badly underestimated Conservative support (Lauderdale, 2015). 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 focus on the sub-sample of of the SoMA panel 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, Panel(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, Panel(a) demonstrates that there are sizable difference between the SoMA sample 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.[6] 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 to have voted for Labour and especially the Green party in the 2015 election, as can be seen in Table 2(b). Our sample does tend to systematically under-report support for UKIP, however. Among both samples, the breakdown by country of resident 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.

As a "control" group, we drew respondents from another YouGov sample–the "Nationally Representative" (NR) sample. These respondents were entirely separate from the SoMA group, but received an identical 4-wave panel survey. Because this sample was representative of the UK population, it included a large number of Twitter users, but because we did not have access to their Twitter accounts, we could not include them in our analysis.

Instead, we only include those respondents in the NR sample who did not use Twitter. Below, we perform analyses that use exposure to tweets as an explanatory variable. For these NR respondents, we assume that there were exposed to 0 tweets. Including these respondents thus allows us to track changes in political knowledge

---

[6]There might be a concern that these median values mask some over-representation of particular demographics, especially young or wealthy people. However, only 10% of our sample is under 30, and only 4% reported a household income over £100,000.

13

among non-Twitter users. Overall, there were 389 NR non-Twitter users who appeared in both waves 1 and 4 (for the party placement analysis below), and 632 who appeared in both waves 2 and 3 (for the factual question analysis).

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,[7] 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.[8] 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.[9] Self-reported measures of media use are fraught with

---

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

[8]Twitter added a "while you were away" feature to highlight tweets that its algorithm predicts the user is likely to be interested in on January 21, 2015, but this represents a tiny fraction of the overall Twitter feed.

[9]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

measurement error (Prior, 2013). Although we ask respondents outright how often they use Twitter, the validity of this information is difficult to verify. We use this variable as a covariate in our analyses, but hesitate to use it to make assumptions about our independent tweet count variables.[10]

To determine which tweets were politically relevant, we manually constructed short lists of terms related to our topics of interest. These topics include UK taxing/spending policy, the UK's ties to EU, legal immigration to the UK, and the extent of ISIS' expansion. From these short lists of "anchor terms" we then identified which other terms most frequently co-occurred with the original terms. We then use these expanded list of terms to determine to identify tweets related to each topics. For a full list of terms, see Appendix A.

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."[11]

_____

perspective) is reasonable as a first step.

[10]We re-did our main analysis restricted to the subset of respondents who claimed to use Twitter "Every few weeks" or more often; the results are not substantively changed.

[11]The advantage of this approach – as opposed to just coming up with our own longer list originally – is two-fold. First, it allows the data itself inform us about the correct terms to use

[**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 split the political accounts into those associated with each of the four major political parties under study. For media accounts, a research assistant identified the UK media organizations with the greatest number of Twitter accounts–including the accounts of journalists employed by those organizations–and we then divided them according to their ideological leanings. Major left-leaning media outlets are The Guardian and The Independent; right-leaning media outlets are The Times and The Sun; centrist media outlets are Scottish TV, the BBC, CNN and The Financial 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 respondents' timelines is shown in Table 4.

[**Table 4 Here**]

Each cell of Panel A shows the number of respondents 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 respondents: we see that Labour and the Conservatives, the two largest parties, have tweets that reach the most respondents, and that centrist media reaches the most respondents overall.

---

in the list, which is especially valuable when using social media where language use is constantly evolving. In addition, the method is replicable: conditional on using the state start words, the algorithm should always produce the same list of 25 most commonly co-occuring words.

Panel B restricts its summary to only those respondents who have seen at least one tweet from a particular source. The first row of Panel B, for example, looks at all of the tweets by Labour to appear in each respondents' timeline, and breaks them down by topic. Among those people who received at least one tweet from Labour, the mean percentage of the Labour tweets about economic issues in their timeline was 49%.

Comparing the rows in Panel B, there is a marked difference in the relative emphases placed on the four topics by each source. For example, nearly half of tweets sent by Labour or the Tories were about the economy, while UKIP tweeted about the economy much less than about immigration or the EU. There is less variation within the media accounts, although the Left Media tended to avoid discussing immigration. On average, media accounts were more likely to tweet about ISIS than were the parties.

Essentially, Panel A describes the breadth of the 28 (7 blocs × 4 topics) information treatments under study, and Panel B describes the relative 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 discussed by the accounts that our respondents have chosen to follow can vary independently, 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: UK taxing/spending policy, the

UK's ties to EU, and legal immigration to the UK. In each wave of the survey[12], 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 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.

All of these placement estimates were highly correlated with each other at .93 or higher, and we use the simplest measure–the mean of each placement–as our "ground truth."[13] 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 CHES was conducted, suggesting that differences in later waves could be 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

---

[12]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.

[13]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.

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 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 issues, but not for the topic of the EU: the Liberal Democrats moved to the right, making their position similar to that of Labour. As a result, we coded the respondent's ranking as "correct" if they placed Labour to the left of the Liberal Democrats or vice versa. 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.[14] Overall, ranking the parties on spending was the most difficult, with only 55 percent of respondents in wave 4 answering doing so correctly among those who attempted to answer it in both waves; the N is considerably smaller for this question.

The results of the statistical tests of H_1 are presented in Figure 4. Each of the three horizontal lines is the logit coefficient (with standard errors) of the logged number of tweets in the respondent's timeline related to the that topic between waves 1 and 4.[15] The dependent variable in each regression is whether the respondent correctly ranked the four parties on that topic in wave 4 of the survey; becaus e this is a binary variable, it is estimated with a logit model. In order to estimate the *change* in knowledge, we control for whether they correctly ranked the four parties on that topic in wave 1. Each regression includes a number of

---

[14]This convergence makes interpreting the "improvement" in ranking the parties on this issue difficult–if someone were to entirely ignore political news for eight months and rank Labour to the left of the Liberal Democrats in both wave 1 and wave 4, our coding strategy considers their political knowledge to have increased. There is no easy solution to this problem, but it should be kept in mind when considering the results.

[15]Throughout the analysis, we use the log of (1.0001 plus) the number of tweets in respondent'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.

other demographic control variables.[16]

We see in Figure 4 that all three of the effects are positive, and that 2 are significant at $p < .05$, while the effect on ranking the parties on immigration is just shy of significant at $p < .10$. These findings support H_1.

[**Figure 4**] here

To get a sense of the magnitude of the effect sizes and the distribution of the independent variables, Figure 5 plots the distribution of relevant tweets on the x axis against the predicted probability that the respondent correctly ranked the parties on that topic in wave 4. This approach sets all other independent variables to their mean values. The general effect is positive, although decreasing density of the tweet count variable on the upper end of the distribution means that at no point do the 95% confidence intervals fail to overlap. The slope of this effect is steepest for ranking the parties on spending: if the typical respondent was exposed to e^{10} tweets about immigration instead of 0, their predicted probability of ranking the parties correctly increases from .56 to .67.

[**Figure 5**] here

To test H_2, we disaggregate these topical tweets by the type of source–if the tweets were sent by a political party (or related politician) or a media organization (or affiliated journalist.) Figure 6 demonstrates that the analysis of Figure 4 is driven by tweets from the parties, as these independently have a positive and significant effect on the topics of spending and the EU (but not immigration). Tweets from the media do not have a significant effect on correctly ranking the parties. These findings lend credence to H_2.

[**Figure 6**] here

---

[16]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.

5.2   Factual Knowledge

The other way we operationalize political knowledge is through factual questions about politically relevant 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?[17]

- (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. Panel A restricts the sample to those people who use Twitter at least "Every few weeks," while Panel B only includes respondents who use Twitter "Less often" or "Never." Overall, there is little difference in either the absolute levels of knowledge or in changes in knowledge between the two samples.

ISIS was empirically the easiest question, and immigration the most difficult. Overall, the rate of *unlearning* the correct answer is similar to the rate of learning the correct answer–compare the bottom left and top right cell of each 2 by 2 box. Put another way, about as many people got the question right in wave 2 but wrong in wave 3 as vice versa. This allows us to explore both the types of tweets that inform and the types of tweets that confuse.[18]

---

[17]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.

[18]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, respondents were able to select "Don't Know" instead of answering the question, so the rate of true guessing should be low.

[**Table 5 Here**]

Our analysis here uses the same specification as for the party placements. We first test H_1 by running 3 logistic regressions where the dependent variable is whether the respondent correctly answered that question in wave 3 and the independent variable is the total number of tweets related to that topic that appeared in their feed between wave 2 and wave 3. Again, we control for demographics, media use and whether they correctly answered the question in wave 2. Figure 7 plots three horizontal lines that represent the logit coefficient (with standard errors) of these primary independent variables.[19]

[**Figure 7 Here**]

Figure 7 shows a positive and significant result for the question about immigration, but not for the questions about unemployment and ISIS, lending weak support to H_1. However, the standard errors for the estimate on ISIS are quite large. This is due to the comparatively few tweets on that topic (see Table 4, Panel B). The estimate is positive, but small: 88% of respondents got the question correct in both wave 2 and wave 3, so there is little change in knowledge to be explained.

We again disaggregate these tweets by their source, to test the hypothesis (H_3) that only media tweets, not political party tweets, will increase factual knowledge. The results are plotted in Figure 8, and concord with our expectations. For all three topics, the effect of media tweets (but not political party tweets) is positive and significant. In fact, on the question of unemployment, the effect of tweets from political parties is *negative* and significant, a finding we explore more below.

[**Figure 8 Here**]

We expect that some political parties have strategic incentives to communicate about some topics with different framings, and that these framings might have differential effects on change in knowledge, our H_4. In particular, the incumbent parties (the Conservatives and Liberal Democrats) should want to emphasize the

---

[19]For space reasons, we do not include effect-size plots for each of these regressions in the body of the text, but see Appendix B.

fact that the UK economy was performing particularly well, while the opposition parties (especially Labour, as UKIP was more focused on non-economic issues) should criticize the economy and make people less likely to believe the truth that unemployment had been going down. Also, fears over immigration were central to UKIP's platform, so they were likely to frame their discussion of immigration to make it seem that the number of immigrants is high.

To test these ideas, we further disaggregate the political party tweets by party (Labour, Conservatives, Liberal Democrats and UKIP) and the media tweets by ideological leanings (Left, Center and Right). We perform the same analysis as above, but now with 7 independent tweet count variables. The results are shown in Figure 9.[20]

[**Figure 9 Here**]

The results of our analysis of the Unemployment question largely agree with our expectations. Tweets from Labour as associated with a *negative* and significant change in knowledge about the unemployment rate, while tweets from the Conservatives have a positive effect, although it falls just shy of significance at $p < .1$. Tweets from Liberal Democrats have a slightly negative effect, contrary to our expectations of a positive effect. Our post-hoc explanation is that the Liberal Democrats suffered greatly from their alliance with the Conservatives, in which they helped support an increase in tuition fees that enraged their constituency. As a result, they actually tried to distance themselves from the coalition government, which would explain our contrary findings.

We also see that the positive effect of media tweets on unemployment knowledge observed in Figure 8 was largest among tweets sent by right-leaning media accounts. A plausible explanation is that the these outlets were sympathetic to the Conservative government, and emphasized the decreasing unemployment rate.

---

[20]We do not plot the results for the analysis of the ISIS question. We have no theoretical expectation about which source should have the largest impact on knowledge of ISIS, and there are so few tweets about ISIS relative to the other topics that the standard errors of our estimates of the disaggregated tweet counts are large.

However, we fail to find the expected positive effect of UKIP tweets on knowledge of immigration. The only party's tweets to be significantly associated with an increase in knowledge of immigration is Labour, and this association is in fact positive. This finding fails to support H_4.

To explain these results, we take advantage of the fact that the multiple choice questions used to measure factual political knowledge about immigration and unemployment had ordinal choices. Instead of merely analyzing changes in *correctness*, we can look at the *direction* of those changes. We fit an ordered probit model where the dependent variable is the difference in the respondent's answer to the relevant multiple choice question. For example, this dependent variable takes a value of 2 if the respondent's answer to the question about immigration went from "Between 100,000 and 300,000 [immigrants] per year" (the second-lowest category) to "More than 500,000 [immigrants] per year" (the highest category). If the respondent instead changed from "Between 100,000 and 300,000 [immigrants] per year" to "Less than 100,000 [immigrants] per year," the DV takes a value of -1.

[**Table 6 Here**]

Table 6 displays the results of these ordered probit regressions. We use the same suite of demographic and media use controls as in previous analyses.

The results of column 1 give context to the evidence from Figure 9: tweets from Labour increase estimates of the change in the rate of unemployment, and tweets from Conservatives and right-leaning media decrease those estimates. The fact of the matter is that unemployment had been decreasing. Because "Decreased" was the lowest possible response (lower than "Stayed the same" or "Increased"), this implies that Labour's tweets were associated with less accuracy and the Conservatives' with greater accuracy. This further supports H_4.

Column 2 of Table 6 explains the null results from the analysis of immigration knowledge in Figure 9. Labour tweets are significantly associated with a decreased estimate of the rate of immigration, while UKIP tweets are significantly associated with an increased estimate. This is precisely what we would expect, based on the strategic frames most useful to these parties and especially to UKIP. The reason

24

that these changes did not necessarily reflect an increased chance of correctly answering the question is that the correct answer ("Between 100,000 and 300,000 [immigrants] per year") was the second lowest choice. UKIP's tweets caused some respondents to correctly raise their estimates from the lowest to the second lowest choice, but caused others to incorrectly raise their estimates beyond the second lowest choice.

These findings do not concord with H_4 as written, but they are plausibly supportive of the idea that the strategic frames used by parties to describe certain issues can impact respondents' factual knowledge.

Overall, we find moderate support for our hypotheses, and explain the cases in which support is lacking. The primary effect of exposure to tweets related to certain topics is to increase knowledge of those topics. Exposure to political information sent by parties tends to increase knowledge of party platforms but not of factual knowledge, while the inverse is true for information sent by media accounts. Political parties do, however, tend to affect factual knowledge of 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 in knowledge rather than levels, we remove the cross-sectional heterogeneity. Further, by matching survey responses to objective measures of political information on social media, we are able to remove self-reported media usage from our independent variable.[21] The net result is "real-world" evidence of the effect of social media exposure on political knowledge.

Our findings largely bear out our initial expectations, although they do admit

---

[21]Although it should be noted we still utilized self-reported usage for robustness tests to address the inherent uncertainty in knowing whether any particular tweet has actually been seen by any given individual.

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 respondents' timelines they actually saw. 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 respondents, 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.

All that being said, it is worth noting that the worst concerns about the relationship between social media and political knowledge do not seem to be have been realized in the 2015 British election campaign. Yes, exposure to UKIP tweets during the election campaign did lead to an upward revision in an individual's belief in the number of immigrants coming to the UK in the past year, but this did not on balance lead to a decrease in correct knowledge. And indeed, across the whole study, we did not simply witness those who say more political tweets learning more "incorrect" knowledge about politics. In fact, this almost never happened, and the one case in which it did – the main opposition party (Labour)'s tweets obscuring the fact that unemployment had fallen – is the one which decades of research on economic voting suggests should not have surprised us. In their own ways, exposure to tweets about politics generally, tweets from political parties, and even tweets from the media seem to have increased political knowledge in at least some areas some of the time. Thus in the first analysis – to our knowledge – to have combined respondent's twitter feeds with panel survey data over the course of an election campaign, we find evidence consistent with the idea that exposure to politics on Twitter may actually be contributing to a more politically informed mass public.

References

Andersen, Robert, James Tilley and Anthony F Heath. 2005. "Political knowledge and enlightened preferences: party choice through the electoral cycle." *British Journal of Political Science* 35(02):285–302.

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 A Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348(6239):1130–1132.

Barabas, Jason, Jennifer Jerit, William Pollock and Carlisle Rainey. 2014. "The question (s) of political knowledge." *American Political Science Review* 108(04):840–855.

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* p. 0956797615594620.

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.".

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

Chang, Linchiat and Jon A Krosnick. 2009. "National surveys via RDD telephone interviewing versus the internet comparing sample representativeness and response quality." *Public Opinion Quarterly* 73(4):641–678.

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. 1997. *What Americans know about politics and why it matters.* Yale University Press.

Evans, Geoffrey and Jon Mellon. 2015. "Working Class Votes and Conservative Losses: Solving the UKIP Puzzle." *Parliamentary Affairs* p. gsv005.

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

Flaxman, Seth, Sharad Goel and Justin M Rao. 2013. "Ideological segregation and the effects of social media on news consumption." *Available at SSRN 2363701* .

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.

Iyengar, Shanto, Kyu S Hahn, Heinz Bonfadelli and Mirko Marr. 2009. "Dark Areas of Ignorance Revisited Comparing International Affairs Knowledge in Switzerland and the United States." *Communication Research* 36(3):341–358.

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.

Lauderdale, Ben. 2015. "What We Got Wrong In Our 2015 U.K. General Election Model." *fivethirtyeight.com* .

Lewis-Beck, Michael S and Martin Paldam. 2000. "Economic voting: an introduction." *Electoral studies* 19(2):113–121.

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.

Prior, Markus. 2005. "News vs. entertainment: How increasing media choice widens gaps in political knowledge and turnout." *American Journal of Political Science* 49(3):577–592.

Prior, Markus. 2007. *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Prior, Markus. 2009. "Improving media effects research through better measurement of news exposure." *The Journal of Politics* 71(03):893–908.

Prior, Markus. 2012. "Who watches presidential debates? Measurement problems in campaign effects research." *Public Opinion Quarterly* 76(2):350–363.

Prior, Markus. 2013. "The challenge of measuring media exposure: Reply to Dilliplane, Goldman, and Mutz." *Political Communication* 30(4):620–634.

Prior, Markus and Arthur Lupia. 2008. "Money, time, and political knowledge: Distinguishing quick recall and political learning skills." *American Journal of Political Science* 52(1):169–183.

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.

Scharkow, Michael. 2016. "The Accuracy of Self-Reported Internet UseA Validation Study Using Client Log Data." *Communication Methods and Measures* 10(1):13–27.

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.

Weaver, Matthew. 2015. "Nick Clegg's tuition fees 'debacle' undermined trust, says Norman Lamb." *The Guardian* May 12.

Whiteley, Paul, Harold D Clarke, David Sanders and Marianne C Stewart. 2013. *Affluence, Austerity and Electoral Change in Britain.* Cambridge University Press.

Table 1: Retention Rates Among Survey Respondents

	Wave 1	Wave 2	Wave 3	Wave 4	All 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)
New respondents		32%	19%	0%	

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 nationally 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	Census
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 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

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 Receiving At Least One Apropos Tweet

	ISIS	EU	Economy	Immigration	Any	Any %
Labour	522	532	532	532	532	19%
Tory	439	470	472	470	472	17%
LibDem	201	223	224	224	224	8%
UKIP	102	102	102	102	102	4%
Right Media	177	184	184	184	184	6%
Left Media	157	161	160	160	161	6%
Centrist Media	701	763	728	761	763	27%

Panel B: Average % of Tweets On Each Topic, Among Those With At Least One

	ISIS	EU	Economy	Immigration
Labour	3%	15%	**49%**	34%
Tory	3%	25%	**45%**	27%
LibDem	1%	29%	42%	28%
UKIP	1%	**36%**	19%	**44%**
Right Media	4%	25%	38%	33%
Left Media	6%	33%	35%	25%
Centrist Media	6%	26%	35%	33%

Table 4: Panel A summarizes how many respondents 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 522 respondents. There are 2,755 respondents for whom we have timeline access. Panel B displays the mean percentage of tweets about each topic sent by each source, of those respondents who saw at least one tweet from that source. For example, the bottom right corner says that, among the 763 respondents with at least one tweet sent by Centrist Media in their timeline, the mean percentage of tweets about immigration–among the tweets sent by Centrist Media about one of the four topics under study–in their timelines is 33%. Cells bolded for emphasis.

Panel A: Factual Knowledge Among Twitter Users (N=1,325)

	ISIS		Unemployment		Immigration	
	Correct W2	Incorrect W2	Correct W2	Incorrect W2	Correct W2	Incorrect W2
Correct W3	89%	6%	53%	9%	33%	18%
Incorrect W3	2%	3%	13%	24%	19%	30%
Total W2	91%	9%	66%	33%	52%	48%

Panel B: Factual Knowledge Among Non-Twitter Users (N=1,076)

	ISIS		Unemployment		Immigration	
	Correct W2	Incorrect W2	Correct W2	Incorrect W2	Correct W2	Incorrect W2
Correct W3	85%	6%	50%	12%	33%	14%
Incorrect W3	5%	5%	11%	27%	24%	30%
Total W2	90%	11%	61%	39%	57%	44%

Table 5: Distribution of Responses to Knowledge Questions: Cell entries are percentages for each possible combination of correct and incorrect answers across wave 2 and wave 3 of the knowledge questions: (C,C), (C,I), (I,C), (I,I). The bottom line shows how difficult each question was showing the percentage correct in wave 2. The sample in Panel A is the respondents who answered the factual questions in both wave 2 and wave 3 and who use Twitter at least "Every few weeks," while Panel B are the respondents who use Twitter "Less often" or "Never."

Effect of Tweets on Estimates of Absolute Levels of Unemployment/Immigration

	Dependent variable:	
	Estimate of Unemployment W3 - Estimate of Unemployment W2	Estimate of Immigration W3 - Estimate of Immigration W2
Labour Tweets	0.091**	−0.040[†]
	(0.020)	(0.021)
UKIP Tweets	-0.008	0.074*
	(0.041)	(0.030)
LibDem Tweets	0.014	−0.045
	(0.034)	(0.037)
Tory Tweets	−0.044	− 0.001
	(0.029)	(0.029)
Right Media Tweets	−0.31	−0.007
	(0.051)	(0.042)
Center Media Tweets	− 0.033	0.017
	(0.025)	(0.024)
Left Media Tweets	−0.068	0.086
	(0.049)	(0.055)
Demographic controls	✓	✓
Media Use controls	✓	✓
Observations	1,713	1,398

Note: †p<0.1; *p<0.05; **p<0.01

Table 6: Estimates of the impact of the number of tweets in the respondent's time-line sent by an account affiliated with that party or group of media outlets *and* related to the that topic, calculated from two separate regressions. The dependent variable in each case is an ordinal variable that corresponds to the answer the respondent 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 as missing, but results are robust to coding "Don't Know" to the median category. Each regression includes demographic and media consumption control variables, as well as a control for the response of the respondent in wave 2.

Placement of Parties in Waves 1 and 4 Among Twitter Users

	EU, N= 1,035	
Correct Order W1	LibDem < Labour < Conservatives < UKIP	
Correct Order W4	LibDem = Labour < Conservatives < UKIP	
	Correct W1	Incorrect W1
Correct W4	54%	27%
Incorrect W4	4%	15%
Total W1	58%	42%
	Immigration, N= 1,013	
Correct Order W1	Labour = LibDem < Conservatives < UKIP	
Correct Order W4	Labour = LibDem < Conservatives < UKIP	
	Correct W1	Incorrect W1
Correct W4	64%	14%
Incorrect W4	10%	12%
Total W1	74%	26%
	Spending, N= 798	
Correct Order W1	Labour < LibDem < Conservatives = UKIP	
Correct Order W4	Labour < LibDem < Conservatives = UKIP	
	Correct W1	Incorrect W1
Correct W4	38%	18%
Incorrect W4	16%	27%
Total W1	54%	45%

Placement of Parties in Waves 1 and 4 Among Non-Twitter Users

	EU, N= 471	
	Correct W1	Incorrect W1
Correct W4	39%	29%
Incorrect W4	8%	24%
Total W1	47%	53%
	Immigration, N= 455	
	Correct W1	Incorrect W1
Correct W4	49%	17%
Incorrect W4	11%	22%
Total W1	60%	39%
	Spending, N= 343	
	Correct W1	Incorrect W1
Correct W4	25%	20%
Incorrect W4	16%	38%
Total W1	41%	58%

Table 7: Cell entries are percentages for each possible combination of correct and incorrect answers across wave 1 and wave 4 of the party placement questions: (C,C), (C,I), (I,C), (I,I). The bottom line shows how difficult each question was showing the percentage correct in wave 1. The sample in Panel A is the respondents who answered the party placement questions in both wave 1 and wave 4 and who use Twitter at least "Every few weeks," while Panel B are the respondents who use Twitter "Less often" or "Never." 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.

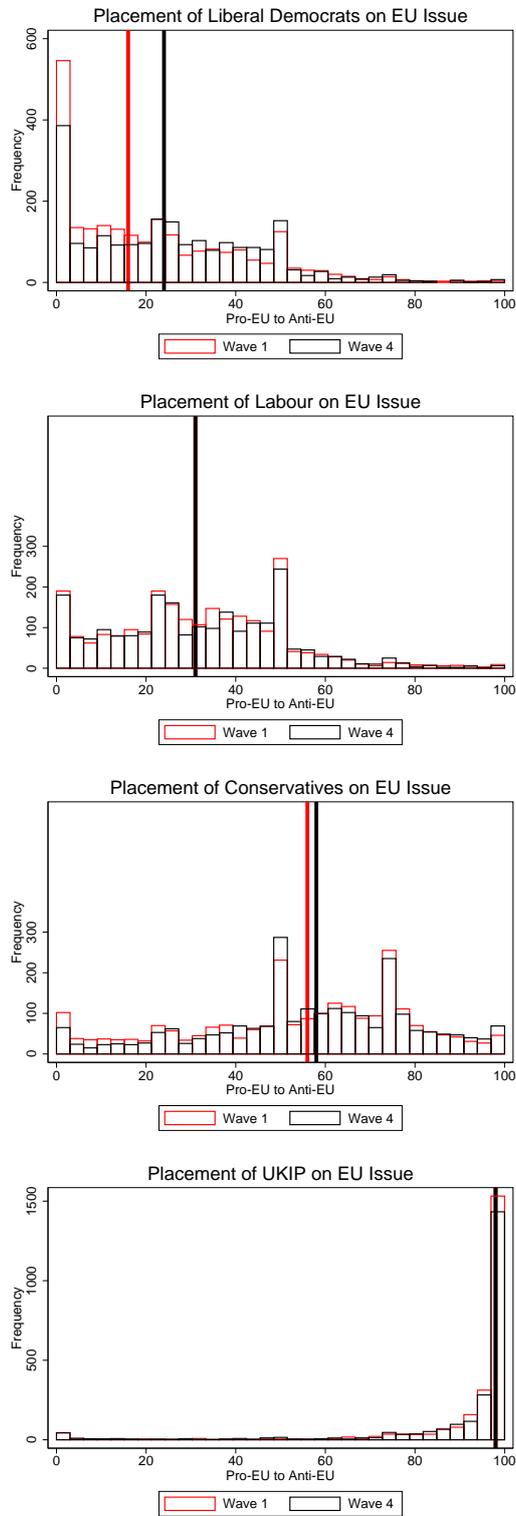Figure 1: Party Placement, Wave 1 to Wave 4: EU
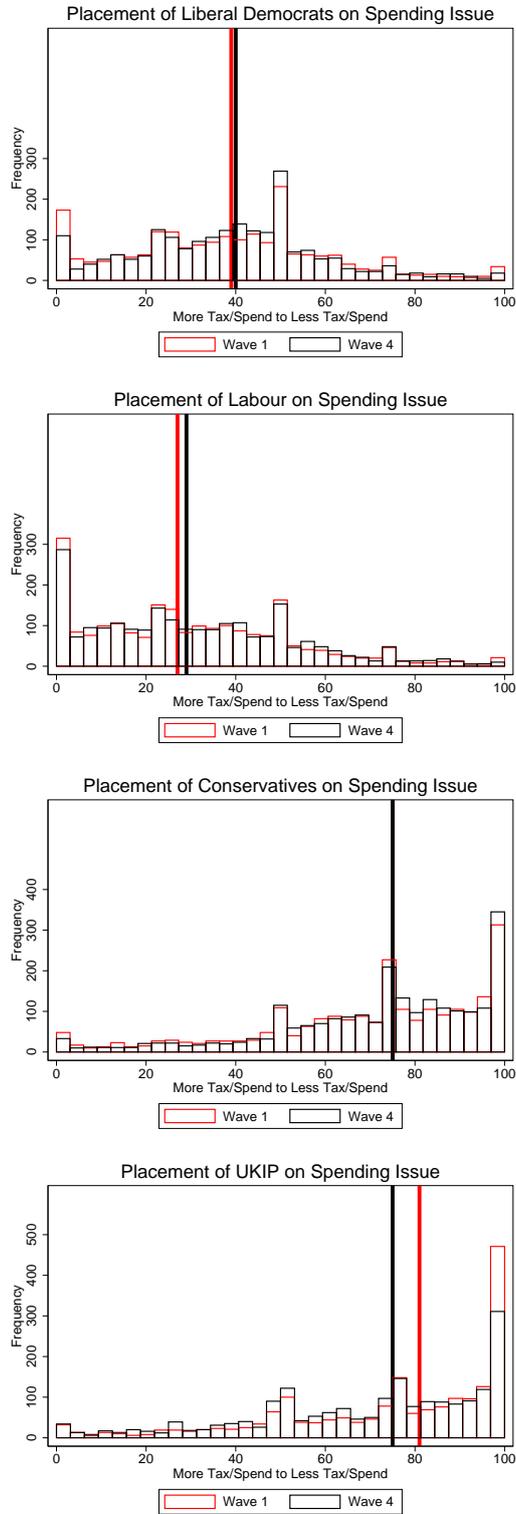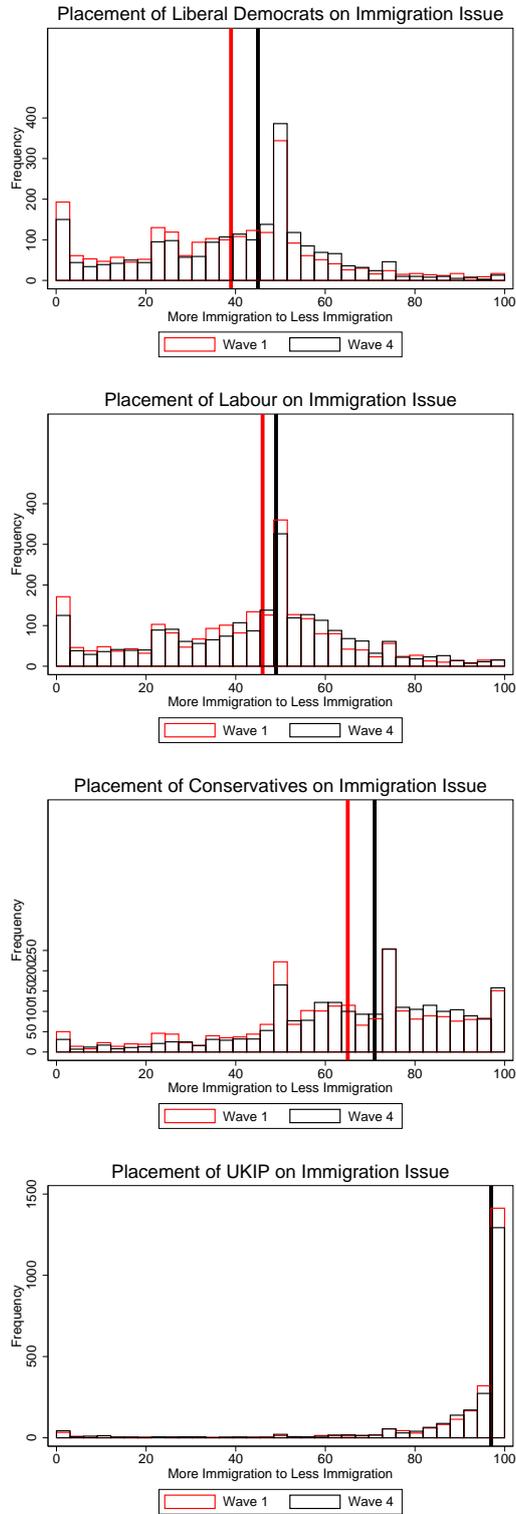
Figure 2: Party Placement, Wave 1 to Wave 4: Spending

Placement of Liberal Democrats on Spending Issue

Placement of Labour on Spending Issue

Placement of Conservatives on Spending Issue

Placement of UKIP on Spending Issue

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

Placement of Liberal Democrats on Immigration Issue

Frequency

More Immigration to Less Immigration

Wave 1   Wave 4

Placement of Labour on Immigration Issue

Frequency

More Immigration to Less Immigration

Wave 1   Wave 4

Placement of Conservatives on Immigration Issue

Frequency

More Immigration to Less Immigration

Wave 1   Wave 4

Placement of UKIP on Immigration Issue

Frequency

More Immigration to Less Immigration
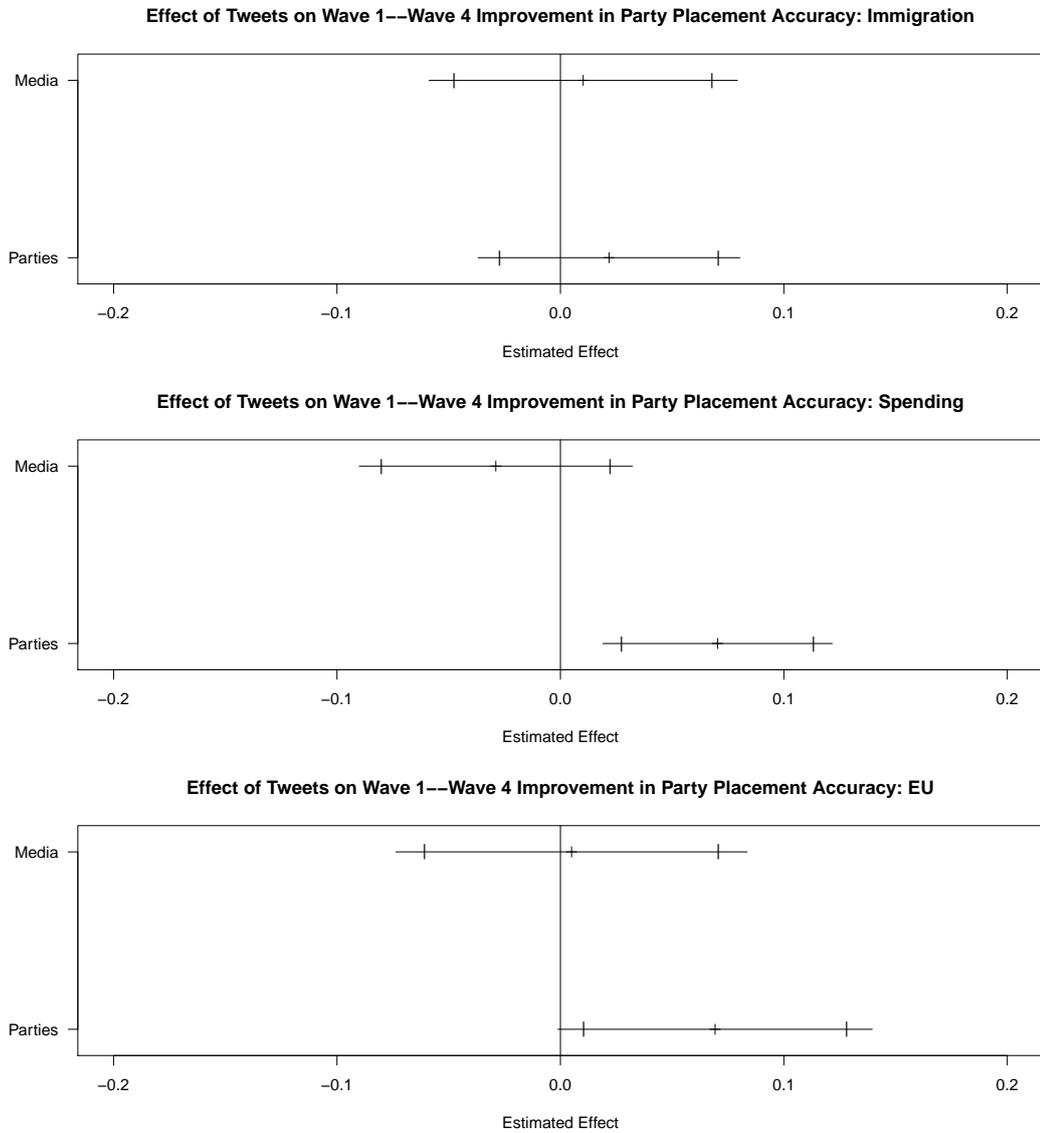
Wave 1   Wave 4

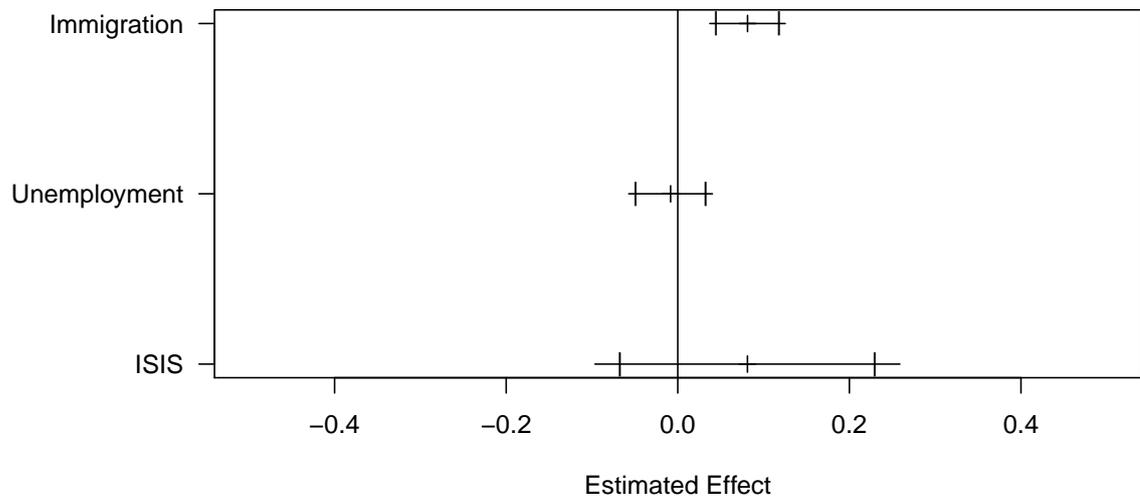**Wave 1--Wave 4 Improvement in Placement Accuracy**
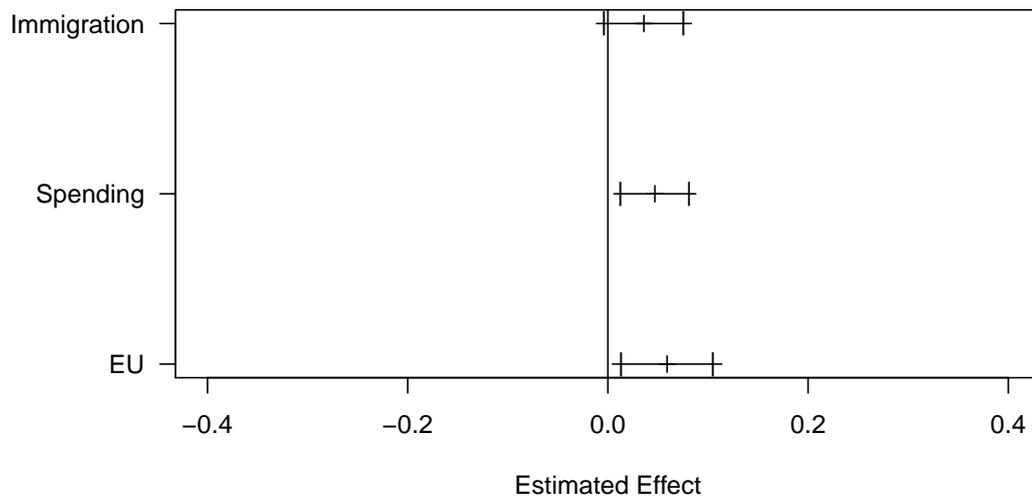
Figure 4: Estimates of the impact of the number of tweets in the respondent's timeline related to the that topic, with three separate regressions. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the respondent correctly ranked the four parties on that topic in wave 4 of the survey; because this is binary, it is estimated with a logit model. Each regression includes demographic and media consumption control variables, as well as a control for whether the respondent correctly ranked the parties on that topic in wave 1.
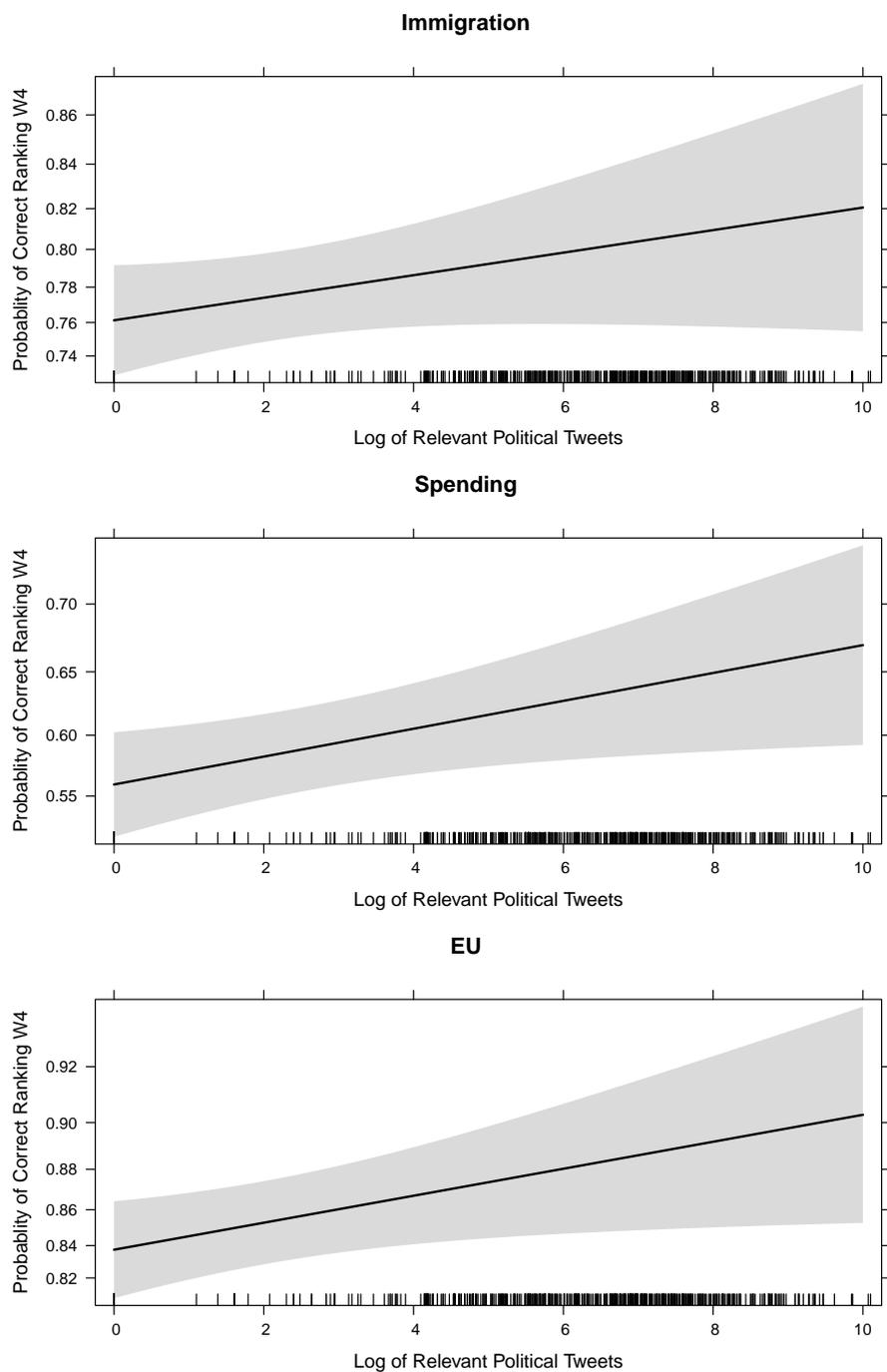
Figure 5: Effects plot of the impact of the number of tweets in the respondent's timeline related to the that topic on the probability that they correctly ranked the four parties on that topic in wave 4 of the survey. This assumes that all other variables, including demographic and media consumption control variables, are fixed at their means.
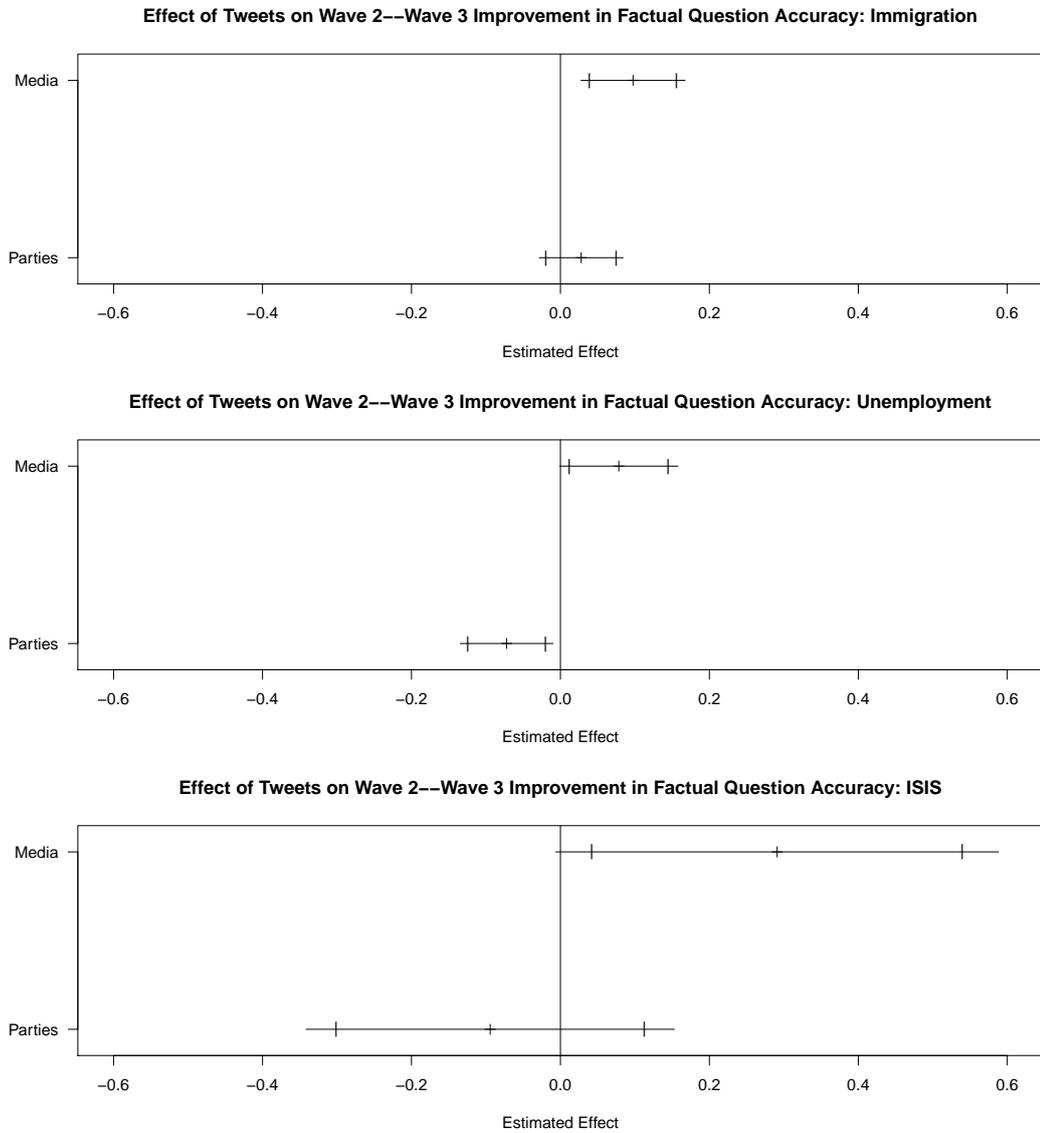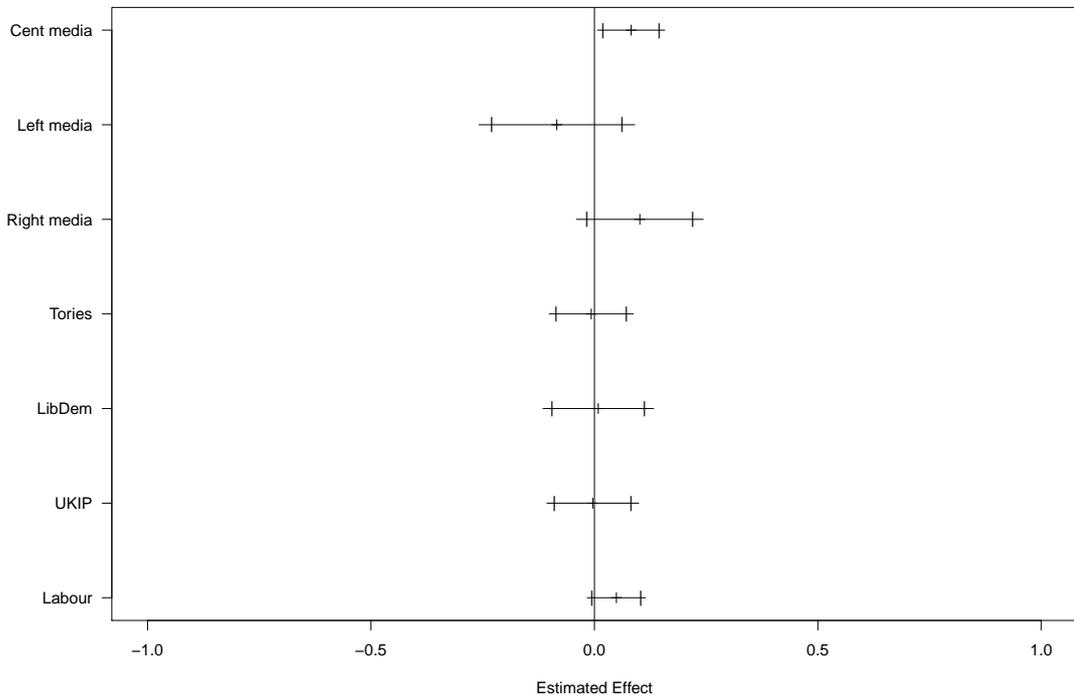
Figure 6: Estimates of the impact of the number of tweets in the respondent's timeline related to the that topic disaggregated by source, with three separate regressions. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the respondent correctly ranked the four parties on that topic in wave 4 of the survey; because this is binary, it is estimated with a logit model. Each regression includes demographic and media consumption control variables, as well as a control for whether the respondent correctly ranked the parties on that topic in wave 1.

Wave 2––Wave 3 Improvement in Factual Accuracy

Figure 7: Estimates of the impact of the number of tweets in the respondent's timeline related to the that topic, with three separate regressions. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the respondent correctly answered the factual question on that topic in wave 3 of the survey, estimated with a logit model. Each regression includes demographic and media consumption control variables, as well as a control for whether the respondent correctly answered the factual question on that topic in wave 2.

Effect of Tweets on Wave 2−−Wave 3 Improvement in Factual Question Accuracy: Immigration

Effect of Tweets on Wave 2−−Wave 3 Improvement in Factual Question Accuracy: Unemployment

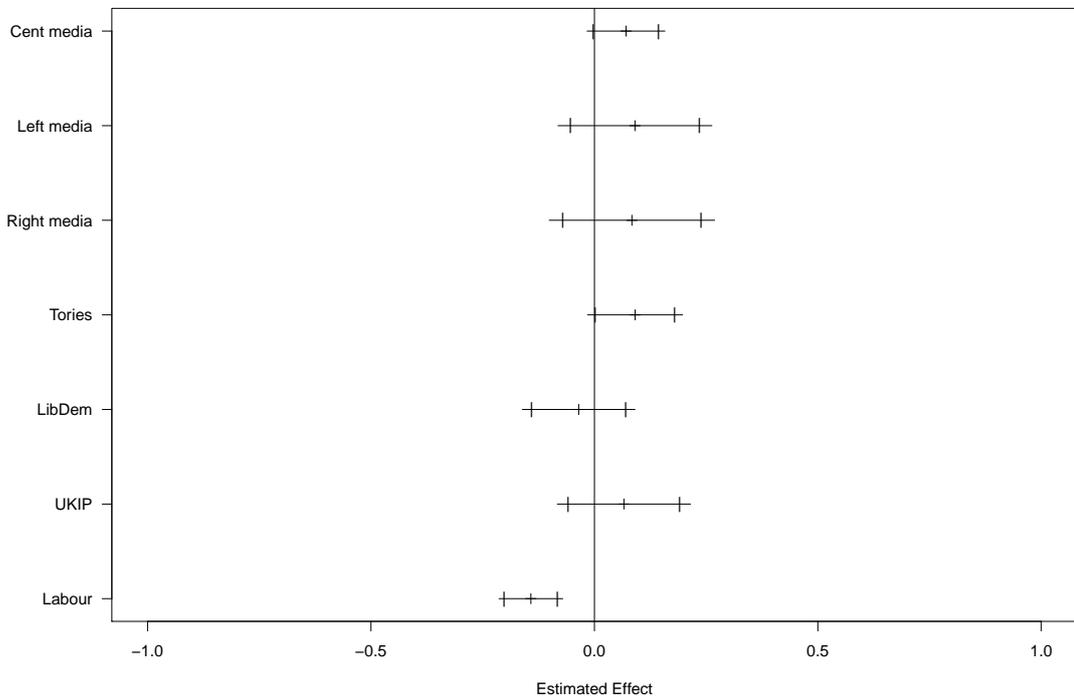**Effect of Tweets on Wave 2−−Wave 3 Improvement in Factual Question Accuracy: ISIS**

Figure 8: Estimates of the impact of the number of tweets in the respondent's timeline related to the that topic disaggregated by source, with three separate regressions. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the respondent correctly answered the factual question on that topic in wave 3 of the survey, estimated with a logit model. Each regression includes demographic and media consumption control variables, as well as a control for whether the respondent correctly answered the factual question on that topic in wave 2.

45

Effect of Tweets on Wave 2––Wave 3 Improvement in Factual Question Accuracy: Immigration

Effect of Tweets on Wave 2––Wave 3 Improvement in Factual Question Accuracy: Unemployment

Figure 9: Estimates of the impact of the number of tweets in the respondent's timeline related to the that topic disaggregated by source, with two separate regressions. Tick marks are 90% confidence intervals; line segments extend to 95% confidence intervals. The dependent variable in each case is whether the respondent correctly answered the factual question on that topic in wave 3 of the survey, estimated with a logit model. Each regression includes demographic and media consumption control variables, as well as a control for whether the respondent correctly answered the factual question on that topic in wave 2.

Appendix A: Terms Used for Topic Creation

The following are the terms used to create each of the topics analyzed in the paper. If a tweet contained terms from multiple topics, it was labeled as belonging to each of those topics.

ECONOMY: cuts benefits budget welfare vat osborne tax tory disabled tories spending austerity cut reform benefit ids nhs ifs labour disability budget2015 health cameron reforms government

ISIS: isis jihad kobane islam iraq syria fundamentalist iraqi mosul kurds kurdish quran ypg raqqa palmyra islamic twitterkurds fighters ramadi muslim kobani beheading bb4sp beheadings peshmerga

UNEMPLOYMENT: unemployment rate muthafukka youth zerohours nsubsides welfarereform lowest figures toryscum falls jobless employment wages underemployment jobsreport jobs nspain psychocrats massaging longtermplan ngreece satire wca unemployed

IMMIGRATION: immigration detention uncontrolled ukip obama farage policy controls reform leadersdebate immigrants illegal eu labour yarl mug bbcqt mass bordersecurity nigel ncustoms time4atimelimit noamnesty debate immigrant

TIES TO THE EU: brexit no2eu betteroffout eureferendum eu euref grexit scoxit stayineu flexcit referendum ciuriak yestoeu ivotedukip nothankeu noxi spexit nunelected efta frexit uk scaremongers anually irexit britty
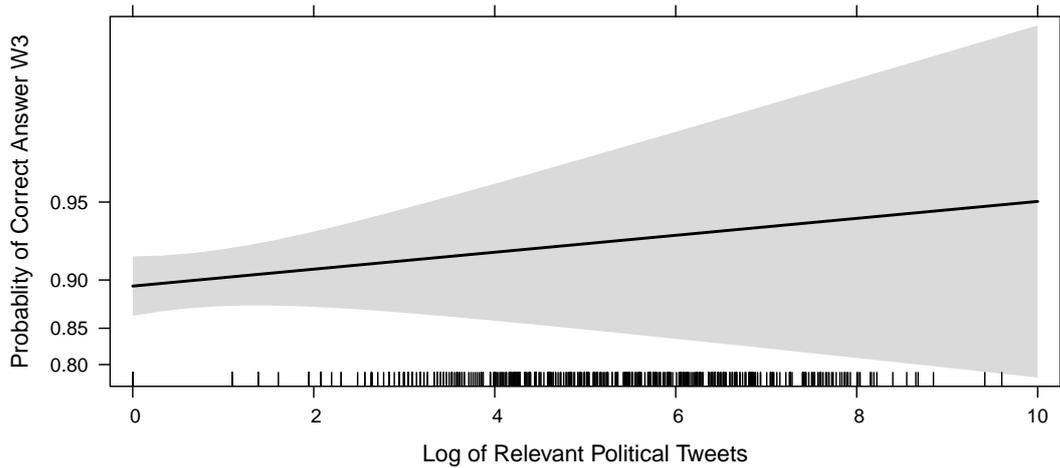
Appendix B: Effects Plots
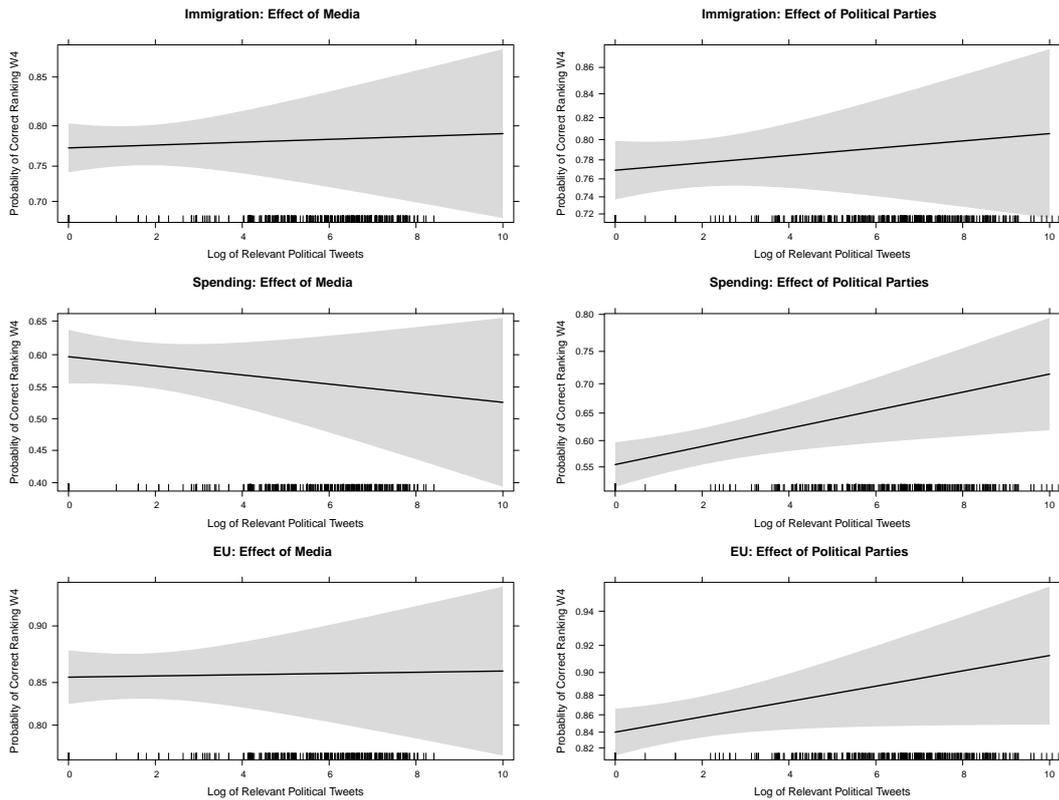
These plots use the same analysis as those in Figure 6. Effects plot of the impact of the number of tweets in the respondent's timeline related to the that topic by parties or the media on the probability that they correctly ranked the four parties on that topic in wave 4 of the survey. This assumes that all other variables, including demographic and media consumption control variables, are fixed at their means.
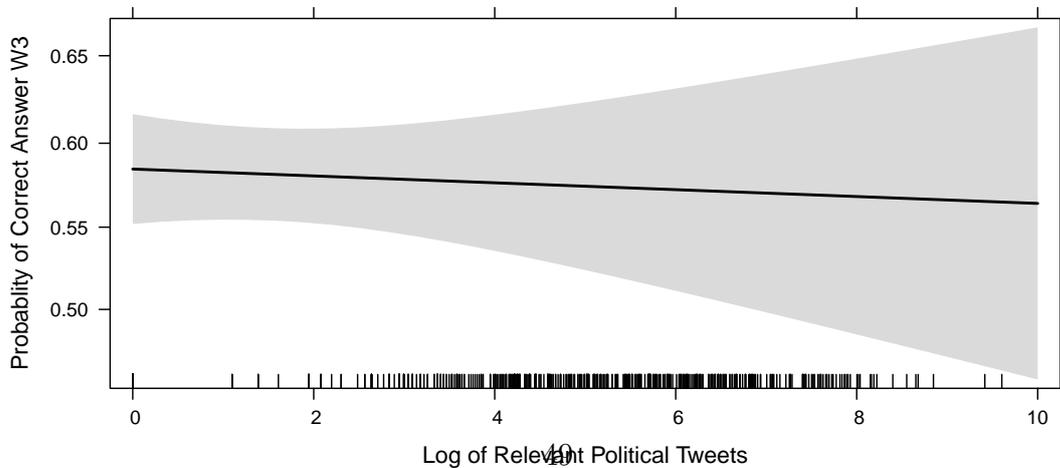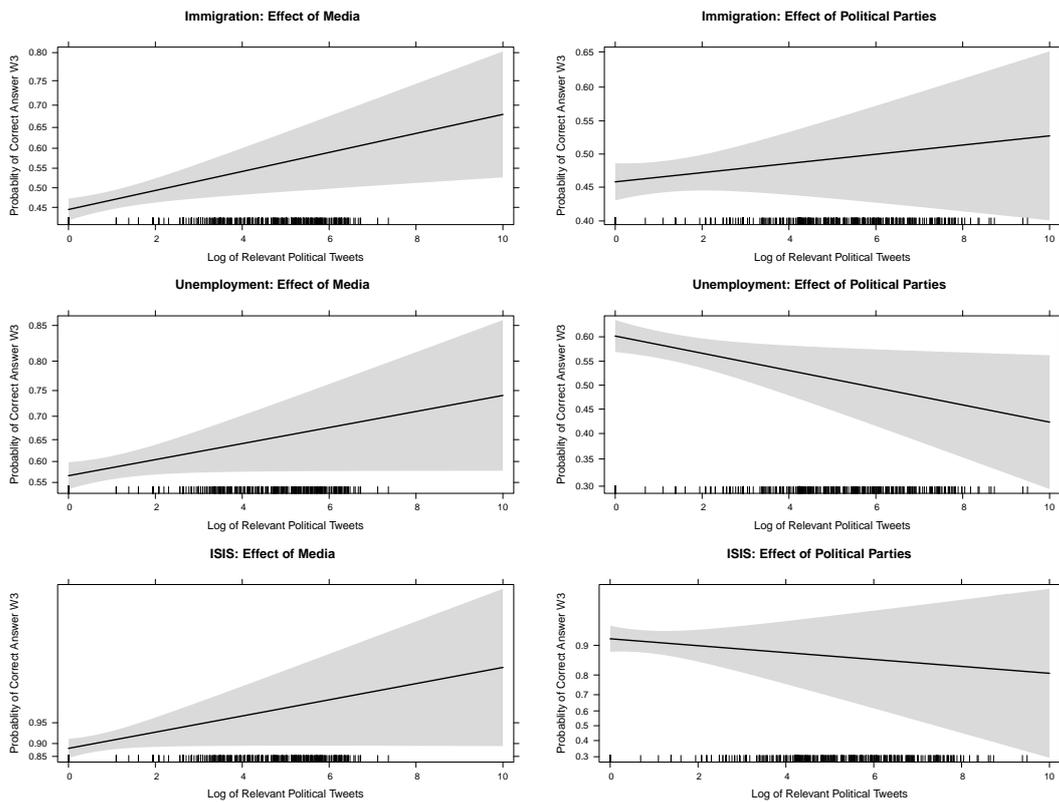
Immigration
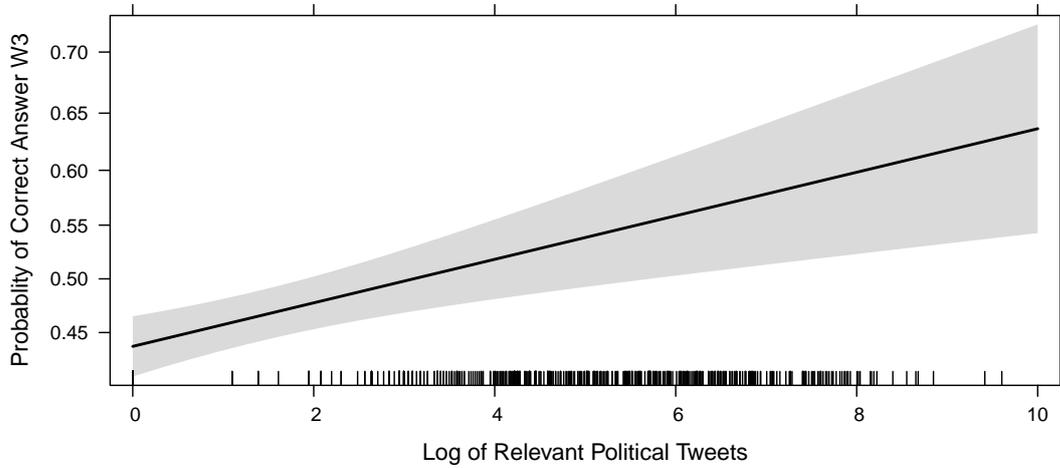
ISIS

Unemployment

These plots use the same analysis as those in Figure 7. Effects plot of the impact of the number of tweets in the respondent's timeline related to the that topic on the probability that they correctly answered the factual question on that topic in wave 3 of the survey. This assumes that all other variables, including demographic and media consumption control variables, are fixed at their means.

These plots use the same analysis as those in Figure 6. Effects plot of the impact of the number of tweets in the respondent's timeline related to the that topic by parties or the media on the probability that they correctly ranked the four parties on that topic in wave 4 of the survey. This assumes that all other variables, including demographic and media consumption control variables, are fixed at their means.