White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/196013/

Version: Accepted Version

Article:
Gorodnichenko, Yuriy, Pham, Tho and Talavera, Oleksandr (2021) Social media, sentiment and public opinions: Evidence from #Brexit and #USElection. European Economic Review. 103772. ISSN: 0014-2921

https://doi.org/10.1016/j.euroecorev.2021.103772

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

SOCIAL MEDIA, SENTIMENT AND PUBLIC OPINIONS: EVIDENCE FROM #BREXIT AND #USELECTION

Yuriy Gorodnichenko
University of California, Berkeley

Tho Pham
University of Reading

Oleksandr Talavera
University of Birmingham

Abstract: This paper studies information diffusion in social media and the potential role of bots in influencing public opinions. Using Twitter data on the 2016 E.U. Referendum ("Brexit") and the 2016 U.S. Presidential Election, we find that diffusion of information on Twitter is largely complete within 1-2 hours. Stronger diffusion between agents with similar beliefs is consistent with the "echo chambers" view of social media. Our results are consistent the notion that bots could have a tangible effect on the tweeting activity of humans and that the degree of bots' influence depends on whether bots provide information consistent with humans' priors. Overall, our results suggest that the aggressive use of Twitter bots, coupled with the fragmentation of social media and the role of sentiment, could enhance political polarization.

JEL classification: D70; D72; L86
Keywords: Brexit; U.S. Election; Information diffusion; Echo chambers; Political Bots; Twitter

* We are grateful to the editors, anonymous referees, Nick Sander and participants of the 2017 Royal Economics Society Conference, the 12th Annual Conference - Warsaw International Economic Meeting, and the 15th Annual Media Economics Workshop for their comments.

1. Introduction

The rise of the internet has changed the way people communicate and acquire information. Consumption of traditional news media has declined in recent years (Stempel et al., 2000) while the number of online news users has soared (Gottfried and Shearer, 2016; Bialik and Matsa, 2017). Among different types of Internet-based media, social networks have become an increasingly important information source for many people. Through social media, individuals can have instant and open access to news or narratives and can build networks to interact and share opinions. Key questions are how this communication revolution has influenced information flows between individuals and how one can influence these flows.

We attempt to answer this question by examining dissemination of information in social media using recent developments in the U.K. (2016 E.U. Referendum, also known as Brexit) and the U.S. (2016 Presidential Election) as two natural experiments. These two events with diametrically opposed political platforms were extremely high-profile so that we would expect people paid considerable attention to these issues. In this exercise, we identify sources of information that might have been used to shape public opinions. Specifically, we study two types of social media agents: real ("human") users and social bots, computer algorithms used to produce automated content. Bots can be used to target specific political messages—either news or opinion—to very specific groups of people or to communicate a (potentially false) sense of consensus on an issue or political candidate. We use these bots as a source of variation in people's information sets and see how this information influences "humans", how it is spread across "humans", and how the sentiment (tonality) of bots' messages affects "humans". While using observational data from these two events does not allow us to establish causal effects, the richness of these two natural experiments permits us to document suggestive evidence on the matter.

Our analysis uses data from Twitter, one of the most popular microblogging platforms with a large number of users. For example, as of 2016, the number of U.K. Twitter users is estimated at 15.8 million while the number of U.S. Twitter users is about 67 million (Benoit, 2017). Given this popularity, Twitter generates an enormous quantity of legally available data for research. These data include records of users, tweets, and metadata that allow us to track tweeting activities of different types of Twitter agents (bots and humans). Additionally, compared to other social network sites, Twitter connections are more about connecting with people for information sharing purposes rather than personal interactions (Gruzd et al., 2011). Hence, one would expect to see

information flows throughout all of Twitter as well as different information clusters during important political events. Furthermore, some of Twitter's users are people with extreme views whose voices are often invisible in the mainstream media. Thus, data from Twitter could give us a broader picture about public opinions during the E.U. Referendum in the U.K. and the 2016 U.S. Presidential Election beyond what we can observe from the traditional media.

We find that tweets about the Brexit and the 2016 U.S. Presidential Election are disseminated and absorbed among Twitter users within 50-70 minutes. This suggests that information rigidity could be very low for critically important issues with wide coverage or that news cycles in social media are short-lived. We also observe the differential intensity of tweeting activities by user type. For example, "remain" supporters in the Brexit Referendum respond stronger and faster to messages created by other "remain" supporters when compared with the reaction to messages from "leave" supporters.

An interesting feature of Twitter is the ease with which bots can operate on the platform. This suggests that human tweeting activity could be influenced by bots. We expect that the degree of influence depends on whether a bot provides information consistent with the priors of a human. For instance, a tweet by a bot supporting the "leave" campaign is followed by a stronger response of a "leave" supporter than a "remain" supporter. Similarly, Trump supporters' tweets are more likely to increase following messages spread by pro-Trump bots. Further examination shows that the sentiment of tweets could play a potentially important role in how information is disseminated: a message with positive (negative) sentiment generates another message with the same sentiment. These results provide evidence consistent with the "echo chambers" effect in social media; that is, people tend to select themselves into groups of like-minded people so that their beliefs are reinforced while information from outsiders might be ignored. Therefore, social media platforms like Twitter could enhance ideological segmentation and make information more fragmented rather than more uniform across people.

This study is related to several strands of research. The first strand has assessed the influence of access to media and news media on political participation. For example, greater access to broadband Internet has led to an increase in partisan hostility and partisan media consumption, which in its turn leads to increased polarization and lower voter turnout (Falck et al., 2015; Lelkes et al., 2015). Gerber et al. (2009) find that the subscription to either Washington Post or Washington Times increases the support for Democratic candidates. Similarly, DellaVigna and

Kaplan (2007) suggest that the introduction of the Fox News channel improved the vote shares of Republican candidates and increased voter turnout. However, the introduction of television with less political coverage leads to a decline in political knowledge and thus turnout rates (Gentzkow, 2006). Further, there is a variation in the effect of media exposure on political participation across media and groups of viewers. For instance, Prior (2005) suggests the positive effect of increased availability of media choices on political knowledge and turnout among individuals whose preferred content is news. The opposite effect is found for people who prefer entertainment. Velez and Newman (2019) find that political knowledge and turnout decline with the exposure to Spanish-language television.[1] We contribute to this literature by shifting our attention to information flows in online media as well as documenting its correlation with political preferences.

With the development of the Internet, there is an emerging yet limited number of studies that examine the real-life impacts of social media. Steinert-Threlkeld et al. (2015) find a positive correlation between the coordination of messages about Arab Spring on Twitter and the number of protests on the following day. Using a sample of historical tweets related to Egypt's Arab Spring, Acemoglu et al. (2018) show a significant association between Twitter activities indicated by the number of Tahrir hashtags in tweets by opposition leaders and the number of protesters in Tahrir Square. Enikolopov et al. (2019) find that the penetration in VK, a Russian social media service, is positively related to the occurrence of protest activity in 2011 but this relationship is driven by lower coordination costs rather than the spreading of information. Social network sites like Facebook can also serve as a propaganda channel through which hate speech can lead to real-life violent crime (Müller and Schwarz, 2018). Regarding the effects of social media news on economic outcomes, Enikolopov et al. (2018) document that publication of blog posts on corruption of Russian state-owned firms can negatively affect stock prices of the implicated firms. We contribute to this emerging literature by quantifying the contribution of messages targeted to specific groups with the intention of affecting their political beliefs and, thus, potentially real-life voting decisions.

Our study also relates to the literature which examines the motivations of news and information consumption. Mullainathan and Shleifer (2005), Halberstam and Knight (2016), and others suggest that people tend to select themselves into groups of like-minded people, so their beliefs are reinforced. Using Twitter data on the 2012 House candidates and their followers

[1] For more studies on the impacts of media exposure and news media see, e.g., Levendusky (2013), DellaVigna and La Ferrara (2015), Druckman et al. (2018), Peisakhin and Rozenas (2018).

(voters), Halberstam and Knight (2016) measure the degree to which Twitter users engage with other users with similar ideology and find a high degree of homophily (e.g., conservative voters tend to follow each other). Furthermore, the majority of retweets of tweets by Democratic/liberal candidates are created by liberal voters and a similar pattern is also observed for tweets by Republican/conservative candidates. Similarly, the high level of homophily during the 2016 U.S. Presidential Election has been also documented (Caetano et al., 2018). The result provides evidence for an individual to prefer to transmit information consistent with their existing beliefs and reflects a mechanism determining each individual's exposure to future information (Adamic and Glance, 2005; Garrett, 2009a; Gruzd and Roy, 2014; Hong and Kim, 2016). However, Gentzkow and Shapiro (2011) suggest that although ideological segregation does exist in both online and offline news consumption, the level of polarization is low. In other words, exposure to information reinforcing individuals' pre-existing views does not shield them from receiving information that they disagree with (Hargittai et al., 2008; Garrett, 2009b; DellaVigna et al., 2014). There is also evidence that polarization is greater for political issues than non-political issues (Barberá et al., 2015). Our contribution to this strand of research is to quantify the extent to which communications by bots are received and distributed around different parts of the Twittersphere.

The fourth strand of research investigates the role of sentiment in the transmission of information (e.g., Heath, 1996; Chen and Lurie, 2013). In political discourse on Twitter, it is observed that tweets containing a high degree of emotionality reach wider readership and are more likely to be disseminated (Kim and Yoo, 2012; Dang-Xuan et al., 2013; Stieglitz and Dang-Xuan, 2013). In addition, people tend to react faster to the information having emotionally negative tone (e.g., Baumeister et al., 2001). In a similar vein, Meffert et al. (2006) find the existence of negativity bias in voters' information selection, processing, and recall during a political campaign. However, it is not always the case. During the 2012 U.S. General Election, voters were more exposed to the content mentioning their favorite candidates with positive sentiment (Halberstam and Knight, 2016). In a contribution to this line of research, we show how sentiment intensity of messages posted by bots predicts reactions of "human" users.

Finally, this paper contributes to the literature on the presence of bots in social media. It has been shown that social bots become increasingly sophisticated and can mimic human behavior, making bots detection more difficult (Ferrara et al., 2016; Haustein et al., 2016). A recent study by Howard et al. (2018) provides a detailed discussion on the existence and creation of political bots as

well as cases when bots are used to manipulate public opinions and/or interfere political discourse. For example, during elections, political bots can create and pitch information for voters who are seeking for political information while elected officials use bots to increase their popularity. Stukal et al. (2017) detect the high presence of Twitter bots' activities in Russian Twittersphere during the 2014-2015 period and find that bots were mainly involved in spreading news stories. Further, several studies (e.g., Bastos and Mercea, 2017; Bessi and Ferrara, 2016; Howard et al., 2016; Howard and Kollanyi, 2016) have attempted to document the role of Twitter bots in the discussions about the Brexit Referendum and the 2016 U.S. Presidential Election. Generally, the findings suggest that Twitter bots effectively spread information during these two events. These studies focus on describing the automated accounts' tweeting activities as well as examining the interactions between bots and humans by mapping their networks. In contrast, we contribute to this literature by studying the behavior of bots and employing econometric techniques to investigate how bots could influence humans.

The rest of this paper is organized as follows. In the next section, we describe the dataset and how we collected data. Section 3 presents our empirical strategy and results. Section 4 concludes and discusses implications.

2. Data

2.1. Data collection and cleaning

The data for analysis were collected using Twitter Streaming APIs. API can be viewed as a tool for collecting data directly from Twitter in real-time. A user sets selection criteria (filters) such as keywords or location and Twitter sends a sample of selected tweets as they happen.[2] Each retrieved tweet contains the plain text of the tweet as well as information about users like user ID (username) and other fields such as date, source, location, friend counts, follower counts, URL, etc.[3]

González-Bailón et al. (2014) and Morstatter et al. (2013) report that the quality of samples increases with more comprehensive coverage (e.g., longer harvested periods to get larger sample

[2] See for details https://developer.twitter.com/en/docs/tweets/filter-realtime/guides/powertrack_rules_and_filtering (Accessed on 24 August 2018).
[3] See Appendix E1 for detailed discussion on the Twitter Streaming API rate limit.

size or well-specified filtering parameters). Consistent with this recommendation, we make API requests to collect tweets that contain the following keywords and leave the connections open to collect as many tweets as possible.[4] The Brexit-related tweets were tracked if they contain the keyword "Brexit". The 2016 U.S. Election-related tweets were collected if they contained the following keywords: "Election2016", "BlackLivesMatter", "CampaignZero", "ClintonEmails", "ImWithHer", "NeverClinton", "FeelTheBern", "CruzCrew", "MakeAmericanGreatAgain", "Trump", "Clinton".[5]

The screening/cleaning process is as follows. First, we process each tweet to extract the relevant content and store in a new tweet content variable. Specifically, we exclude special characters such as link tokens (starting with "http://", "https://", "www.") or user identifier tokens (starting with "@") from the tweet content. Second, we do not include tweets that contained only links or URLs in our analysis.[6] Third, we separate tweets with English as the language description from those with other languages. After these steps, we obtain a sample of nearly 22 million tweets for the E.U. Finally, we adopt the approach proposed by Howard and Kollanyi (2016) and Howard et al. (2016) and define campaign endorsement for each tweet based on the hashtags/keywords specified in Appendix Table A1. To be specific, a tweet is defined as a pro-leave tweet if it (1) contains at least one of the keywords in column 2, row 1 of Appendix Table A1 and (2) does not contain any pro-remain keywords (column 2, row 2). Similarly, a pro-Trump tweet contains at least one keyword listed in column 2, row 3 of Appendix Table A1 and does not contain any pro-Clinton keywords (column 2, row 4). Since we are interested in examining the interactions between supporters of two sides of the debates, any tweets that do not show a clear-cut political endorsement are not considered in our time series analysis. Our estimation sample contains about 2.7 million tweets for the E.U. Referendum and 7.2 million tweets for the 2016 U.S. Presidential Election.[7]

---

[4] Data used in the analysis is restricted to the tweets posted during 24th May – 23rd July 2016 period for the Brexit sample and 8th October 2016 – 8th December 2016 period for the 2016 U.S. Presidential Election sample.

[5] There are two potential concerns related to our collection of tweets for the U.S sample. First, some keywords could be redundant. However, we screened and cleaned the data before analysis to prevent double counting. Furthermore, the keywords used for data collection are different from the keywords used for classifying political endorsement. Also note that we might miss some tweets that used the short form of the selected keywords (e.g., "MAGA" instead of "MakeAmericaGreatAgain"). This could lead to a lower number of pro-Trump tweets in our sample compared to the actual number of tweets and thus, we would be less likely to find any significant effects. However, this would enhance our findings because, if anything, our results would underestimate the effects of pro-Trump tweets.

[6] This criterion effectively removes Twitter accounts of many standard media (e.g., BBC, The Times) because these media typically post only links (URLs) to articles on their Twitter pages.

[7] Appendix E provides additional details on Twitter API and the effect of filters on the estimation sample.

In the next step, we separate original tweets (i.e., the tweets that were created rather than copied) and their retweets. First, we screen each tweet's content in the original dataset and create a new indicator variable which equals 1 if the tweet starts with "RT @" (which means a retweet in the Twitter language; that is, a repost of an original tweet) and 0 otherwise. Next, we extract the content after "@" and before the main text to create a new variable that contains the username of the Twitter account from which the tweet was retweeted. After these steps, we could identify (1) the original tweets and (2) their retweets (if any). An example of this process is provided in Appendix Table A2.

2.2. Identification of bots

Since Twitter bots are run by software, it is extremely difficult to establish who bot-makers are really and how their content is created. Indeed, bots' tweets are automatedly generated and the content of the tweets could be pre-determined or could be links to popular sites, news sites, or sites focused on current events (Wojcik, 2018). Furthermore, Twitter bots become increasingly sophisticated at mimicking humans (Ferrara et al., 2016; Haustein et al., 2016). However, previous research (e.g., Chu et al., 2010; Sivanesh et al., 2013; Cook et al., 2014) documents several patterns that help distinguish bots and human users. First, a human agent tends to be more active in tweeting during the regular workdays and during the daytime while the daily and hourly tweeting activity of a bot agent is even. Second, bots often tweet the same content many times while humans do not. Given the aggressive use of social bots during the political events like election, previous studies also suggest some event-specific criteria to detect bots. For example, bot accounts are more likely to be created just on or about the event announcement date. Further, bot agents could be inactive before and after the event but create mass tweets or retweets on event-specific days and times.

Building on these earlier results, we use the following procedure to classify Twitter accounts into humans and bots. Consider a given Twitter account on a given day. We flag this account/day as a potential bot if *any* of the following conditions is satisfied.

First, Haustein et al. (2016) document that the average number of daily tweets for a bot is about 5 (standard deviation ≈5) while the daily average for humans is about 2 (standard deviation ≈2). Given these differences in the volume of tweeting, our first criterion is an unusually large number of tweets in a day created by an account. For both samples, we choose the threshold of 10 which is about 4 standard deviations above normal human activity.

Second, since the period from 00:00 to 06:00 is often considered as the inactive time for humans, any activity during this time period could be viewed as "abnormal tweeting time". We flag an account as a potential bot if the account creates at least 5 tweets during the abnormal tweeting time on a given. Abnormal time is defined based on British Summer Time for the U.K. or Eastern Time for the U.S.[8]

Third, previous studies in computer science (e.g., Lee et al., 2010; Chu et al., 2012) suggest that one of bots' characteristics is to repeatedly post identical messages. Thus, we flag an account as a potential bot if the number of tweets with the same content per day is 3 or more for both samples.

Fourth, Twitter bots may be created specifically for and therefore shortly before particular events. We define an account as "newly created" if the account is created on or after the Brexit Referendum Announcement (20 February 2016) for the Brexit sample and 15 July 2016 for the 2016 U.S. Election sample which is when the (at-the-time) Republican presumptive nominee Donald Trump announced his vice-presidential running mate. These accounts are flagged as potential bots if they have an unusually high average daily tweet volume relative to what one might have expected for a Twitter account of that age. The chosen threshold is 10.

If an account is flagged as a bot for majority of days (that is, more than 50 percent of days) during its lifetime in the sample, then the account is defined as a bot; otherwise the user is defined as a human. An example of a bot account is "@**2053pam**", which has been suspended by Twitter. As recorded in our data, this account was created on August 24, 2010 and as of May 23, 2016, it had created more than 93,800 tweets (on average, it tweeted about 45 tweets per day!). Another example is an account named "@**helen_1rivera**", which has also been suspended. It was created on June 12, 2016 and after about half a month, it already created more than 4,000 tweets and attracted more than 200 followers.[9] We experimented with variations of these criteria (e.g., a user is defined as a bot if tweeting activities are observed for at least three days and on more than 50 percent of days tweeting activities match all four criteria; raising or lowering the threshold point for each criterion) and we found similar results in our regression analysis.[10]

---

[8] We find similar results if the tweeting time is classified based on Pacific Time for the U.S.

[9] Examples of bot accounts and their tweets as well as current status can be found in Appendix A4.

[10] Our bot classification is likely "conservative" with a low likelihood of mis-classifying an active human user as a bot. For instance, President Donald Trump's Twitter account is not identified as a bot account in our data.

Table 1 presents summary statistics for bot and human accounts in the Brexit and U.S. Election samples. The number of bot accounts is small (0.24% of total accounts observed in the Brexit sample and 0.2% for the U.S. Election sample) but bot accounts generate a lot of traffic (12.4% of all tweets in the Brexit sample and 8% of all tweets in the U.S. Election sample). The average age of bot accounts is generally younger than that of human accounts. This pattern is consistent with the fact that many bots are often created for a particular event. However, we also identify some bots that have existed for a long period of time, which are likely to be used for marketing/spamming purposes or even active in no-event periods.

[Table 1 goes here]

2.3. Validation of bot identification

We perform multiple checks to ensure the validity of our procedure to identify bots. First, we compare our bot definition with bot detection based on an online classification tool called Botometer (formerly BotOrNot).[11] Due to the constraints related to sample size and rate limit (i.e., the number of searches allowed per unit of time) when using Botometer service, we were only able to perform this check for the Brexit sample. We find that our approach and the Botometer approach return similar results for 90% of accounts in our sample.

Second, since 2018, Twitter has made an effort to tackle social media fraud, resulting in the suspension/removal of tens of millions of suspicious accounts ("bots") (New York Times, 2018). When we check the status of accounts identified as bots in our sample, we find that 50 percent of "our" bots have been deactivated or have been suspended/removed by Twitter. When we conduct a similar check on two separate random samples of human accounts, the deactivated/suspension/removal rate is around 20 percent.

Third, one feature that distinguishes bots from genuine users is that bots tend to post messages at higher frequency. Thus, one would expect the higher number of tweets created by bots compared to that created by humans within a short time interval (e.g., within one minute). We observe this pattern when looking at tweeting activities of bots and humans on two days prior to the voting days (Figure 1). Per-minute tweeting pattern of human accounts is relatively stable compared

[11] This tool is developed by researchers from Indiana University and Northeastern University. Botometer tool cannot classify the accounts that have been suspended or deactivated. See Davis et al. (2016) for more details.

to bots' per-minute tweeting pattern: a human account posts only one message per minute while this number is on average 2-3 for a bot account. Note that this observation is not a direct result of our bot identification as the identification criteria are based on daily tweet frequency.

[Figure 1 goes here]

Fourth, one would expect a decrease in the volume of humans' tweets about Brexit/U.S. Election if there are live (non-political) events that are popular enough to shift humans' attention away from political debates. In contrast, bots' tweeting patterns about Brexit and U.S. Election should not be affected by the events since bots' tweeting activities are pre-programmed and hence, cannot react to live events. During the time span of the Brexit sample, the 2016 UEFA European Championship, in which England, Wales, and Northern Ireland teams played, also took place, with England and Wales being in the same group. Thus, undoubtedly, the 2016 European Championship got significant attention from individuals in the U.K., which could distract them from the EU Referendum debates. This is indeed observed during the matches in which England played against Russia and Wales played against Slovakia on 11th June 2016. During these matches, the number of humans' Brexit-related tweets decreased dramatically. By contrast, bots' Brexit-related tweeting pattern largely unchanged during the matches (Figure 2). This observation also suggests that bots are not programmed to react to humans.

[Figure 2 goes here]

In summary, our identification of bots and users are consistent with identification by others and exhibit behaviors consistent with each group.

2.4.Dynamics of Twitter posts

Figure 3 illustrates the evolution in the daily and hourly volumes of Brexit-related tweets for humans (Panels B and D) and bots (Panels A and C). There is a significant increase in the number of tweets created by humans on the 23rd and 24th June 2016. While bots also show more intensity around these dates, the increase is much more modest. Interestingly, the daily volume of pro-leave tweets was always higher than the daily volume of pro-remain tweets. This gap was greatest during the time around the E.U. Referendum day: between 00:00 and 06:00 on the 24th June, the difference in the hourly pro-leave tweets and pro-remain tweets reached its peak of about 10,000 tweets. There is a clear pattern in humans' hourly tweeting volume: humans' accounts are more active

between 6 am and 6 pm and they show considerably smaller intensity in other hours.[12] In contrast, we do not observe any clear pattern in the hour-by-hour tweeting activity of bots.

[Figure 3 goes here]

The time series of 2016 U.S. Election-related tweets are shown in Figure 4. Most of the time, the number of pro-Trump tweets exceeded the daily volume of pro-Clinton tweets. A large increase in pro-Clinton tweets only appeared during the time running up to the Election Day. Specifically, approximately 5 days before and on the Election Day, the number of pro-Clinton tweets soared with the peak of more than 8,000 tweets per hour and was higher than the number of pro-Trump tweets. Comparing the differences in the number of tweets created by the two sides before and after the voting day, we observe a significant reduction in the hour-by-hour gap between two periods. Note that the intensity of tweeting activity declined sharply after the Election Day while in the U.K. the decline was more spread out in time during the post-Referendum period.

[Figure 4 goes here]

2.5. Sentiment of the tweets

Baumeister et al. (2001), Kim and Yoo (2012), Stieglitz and Dang-Xuan (2013), and others show that the intensity of information flows can depend on sentiment (tonality) of messages. To measure the intensity of sentiment, we use TextBlob, a publicly available text-processing tool written in Python, to get a polarity score for each tweet (see Loria, 2018 for TextBlob details).[13]

The sentiment analysis using PatternAnalyzer in TextBlob is done in the following steps. First, the text is divided into constituent words. Second, stop words such as function words (the, that, which, etc.) are filtered out. Third, part-of-speech tagging is performed. In this step, the remaining words after the second step are classified into categories i.e. noun, verb, adjectives, adverb, etc. Finally, words are passed into Sentiment Classifier to determine the polarity/subjectivity scores for each word and for the whole sentence. TextBlob's Sentiment Classifier scores sentiment mostly based on English adjectives and modifiers then averages sentiment of a single word and the text.

[12] Appendix Figure B1 shows how intensity of tweeting activity by humans and bots changes by hour of the day and by day of the week.

[13] To check the robustness of Textblob dictionary, we compare the positiveness of tweets in our sample based on Textblob dictionary and based on social media dictionary developed by Renault (2017). The correlation coefficient is 47%, when either positive or negative sentiment is detected. The results are similar when we use this alternative approach to measure sentiment.

The abovementioned process returns the polarity score bounded between -1 and 1. A score in [-1,0) represents negative sentiment, a score in (0,1] represents positive sentiment, and a score of 0 refers to neutral sentiment.[14] Overall, the volume of emotionally colored tweets was relatively moderate: neutral messages are the most prevalent (the average share is 50% and 61% for the Brexit and U.S. Election samples, respectively). Messages with positive sentiment are the next-most prevalent (the average share is 33% and 25% for the Brexit and U.S. Election samples, respectively). Negative messages are the least frequent (the average share is 17% and 15% for the Brexit and U.S. Election samples, respectively). The distribution of scores is reported in Appendix Figure B2.

Figure 5 shows the daily volume of tweets by sentiment and type of user for the Brexit sample and the 2016 U.S. Presidential Election sample.[15] Daily volumes by sentiment tend to comove strongly. This pattern is observed for both humans and bots. We also find similar results when we focus on the hourly volume of tweets around the voting dates and when we split the sample by the sides of the campaign (see Appendix Figures B3-B5). This co-movement suggests that the distribution of sentiment was approximately constant during both campaigns.

[Figure 5 goes here]

3. Interactions between bots and humans

3.1. Retweeting

Similar to other social media, Twitter allows users to repost existing messages. Typically, reposting (retweeting) a message means that a user wants to spread the message in his or her social circle. Messages with many retweets are often labeled as popular/trending and, as a result, have higher ranking/priority in internet searches. In other words, a message with many retweets is often treated as important. Because retweeting a message generates a link from a user who originated the message to a user who reposted it, we can observe the direction of the information flow. Thus, retweeting provides us with measures of intensity and direction for interaction between different types of users.

To understand our approach, consider the following fictitious example. Suppose an original (i.e., not a copy) tweet supporting the leave campaign appears at 1pm. We compute how many

[14] See Appendix A3 and Appendix Table A3 for examples of sentiment analysis using our data.
[15] We do not consider the "intensity" of sentiment. For instance, tweets with the negative polarity scores are classified as negative sentiment tweets regardless of the levels of the scores.

retweets between 1pm and 1:10pm this tweet generated by human accounts and by bot accounts. Then we count the number of (new) retweets by humans and bots that were generated for the original tweet between 1:10pm and 1:20pm. This procedure is continued at ten-minute intervals for 2 hours after the original tweet appeared. The resulting path provides us with an impulse response function for the original tweet. We repeat this procedure for all original tweets and compute the average impulse response function.

Figure 6 reports these average impulse response functions by type of users who generated original tweets and by type of users who retweeted original messages. Since most tweets generate few or no retweets, we restrict the sample to relatively popular original tweets (i.e. more than 5 retweets, which is above the 75th percentile of the retweet distribution) to have meaningful variation over time.[16] Panels A and B of the figure show that, relative to traffic generated by humans, traffic generated by bots is small. Indeed, the intensity of bots' retweeting is an order of magnitude smaller than retweeting activity of humans. Humans and bots are most active in retweeting in the first 10 minutes right after the time when original tweets are generated. The number of new retweets reduces over time and reaches a stable level within two hours.

[Figure 6 goes here]

To establish which account type is more likely to respond to a tweet of a given account type, we restrict the sample to include only original messages generated by humans (Panels C and D) or only original messages generated by bots (Panels E and F).[17] Note that the reaction of humans to tweets generated by other humans is slightly stronger than humans' traffic reacting to tweets generated by bots (i.e., compare the magnitude of the responses in Panels C and E for the Brexit sample and Panels D and F for the U.S. sample). In contrast, the volume of bots' traffic is equally weak in retweeting messages of humans and messages of other bots. For instance, during the first 10 minutes since humans post original tweets about Brexit, the number of retweets made by humans is significantly higher than that made by bots (40 retweets vs. 2 retweets for every 10 original tweets, respectively).[18]

---

[16] In the Brexit sample, the average number of retweets per original tweet is 0.5 and the standard deviation is around 4. In the U.S. Election sample, these figures are 0.3 and 12, respectively.

[17] Similar patterns are also observed when we (1) separate tweets by sentiment and (2) restrict sample to the most popular original tweets (the ones get more than 20 retweets, which is above the 95th percentile of retweet distribution).

[18] Given the nature of retweeting in Twitter, i.e., one account can retweet a particular tweet only once, the result in Panel C also indicates that on average, within 10 minutes since a human account tweets an original tweet, 4 other human accounts react to it through retweeting while almost no bot account has immediate reaction to it.

To further understand the interaction across accounts through retweeting activity, we examine the extent to which an account repeatedly retweets from another account. For instance, account A is considered having 3 "links" with account B if A retweets the content created by B 3 times. This can be understood as the (conservative) intensive margin of account A' exposure to account B. In the Brexit sample, the average number of links between a human account and a bot account is 2.86 which is slightly larger than the one between two human accounts (1.47). In contrast, the average bots' exposure to another bot is 11 which is significantly larger than the exposure to a human account (2.16). In the U.S. Election sample, the average humans' exposure to a bot account is 2.83 and that to another human account is 2.09. Yet, the average number of bots' links with another bot is about 3 times larger than the average number of bots' links with humans (10 versus 3). These statistics suggest that a human account's exposure to bots is relatively similar to the exposure to other humans while a bot's exposure to other bots exceeds the exposure to human accounts.

The small difference in humans' retweeting in response to original tweets generated by bots and other human agents could be explained by the fact that bots are designed to mimic humans' behavior. More specifically, it has been documented that more and more tweets spread by bots are witty and provocative (Veale et al., 2015), and that thought-provoking and novel content is likely to capture humans' attention (Vosoughi et al., 2018). Furthermore, given the sophisticated design, Twitter bots can also exploit the trending algorithms to get certain topics or information to trend (Woolley and Howard, 2017). In other words, bots could be effective in spreading the "trending" information which in its turn is likely to be retweeted by human agents.

These patterns lead us to three tentative conclusions. First, information flows are most intensive between humans, while information flows between bots are weak. Second, information flows from bots to humans are tangible while information flows from humans to bots are very weak. Third, all reactions tend to be relatively short-lived in the sense that the vast majority of the reaction is completed within two hours.

These results are consistent with the view that humans had little (if any) effect on bots while bots had a tangible effect on humans. These results also suggest that bots are not likely to be successful in persistently moving tweeting activity of humans. The short duration of the response is consistent with the fast news cycles in social media (e.g., Kwak et al., 2010; Yoo et al., 2016) and/or low information rigidity (Coibion and Gorodnichenko, 2012).

3.2. Time series analysis

While the analysis in the previous section is informative, it is focused on reposts of original messages. Obviously, interaction between different types of users may also happen via generation of new messages. In this subsection, we use standard tools of time-series analysis to construct impulse responses of all messages (that is, retweets and new posts) by a type of users to a message generated by a given type of users.

This exercise relies on two key ingredients. First, we build on our earlier findings for retweets and assume that humans can respond contemporaneously to bots while bots do not respond contemporaneously to humans. Second, to ensure that this identifying assumption holds, we use data at 10-minute intervals. Apart from strengthening our identification, this short duration of time intervals allows us to control for low-frequency variation in the volume of tweeting activity (e.g., days closer to the vote date have higher volume than more distant days). To be clear, we do not have causality in the conventional sense (i.e., we do not have a randomized control trial or a quasi-random experiment). Instead, we rely on the time-series (Granger) notion of causality (i.e., if there is a reaction of *Y* to a shock in *X*, then *X* is causal for *Y*).

3.2.1. Econometric specification

To estimate impulse responses flexibly, we use local projections method developed by Jordà (2005). To see how this method works, suppose that we are interested in estimating reactions of humans supporting campaign X to bots advocating campaign X and to bots advocating campaign Y. The method amounts to estimating $h = 0, \dots, H$ regressions of the following type:

$$\begin{aligned}\ln Human_{t+h,d}^{(X)} = {} & \sum_{k=0}^{K} \alpha_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(X)} + \sum_{k=0}^{K} \beta_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(Y)} \\ & + \sum_{k=1}^{K} \gamma_{X,k}^{(h)} \ln Human_{t-k,d'}^{(X)} + \sum_{k=1}^{K} \phi_{X,k}^{(h)} \ln Human_{t-k,d'}^{(Y)} \\ & + \psi_{X,d}^{(h)} + Seasonal_{X,td}^{(h)} + error_{X,td}^{(h)} \end{aligned} \quad (1)$$

where t and h index ten-minute intervals, d indexes the day of a campaign, $Human_{t+h,d}^{(X)}$ is the volume of new tweets generated by all humans supporting campaign X during the $t+h$ ten-minute interval on day d, $Bot_{t-k,d'}^{(C)}$ is the volume of new tweets by all bots supporting campaign $C = \{X, Y\}$ during the $t-k$ ten-minute interval on day d' where $d' = d$ if the $t-k$ interval is on the

same day with t and $d' = d - 1$ if the $t - k$ interval is on the day proceeding t. Because there are considerable variations in tweeting activity across days of the week, we include $Seasonal_{td}^{(h)}$, a set of "seasonal" day-hour dummy variables. Specifically, for each 1-hour interval during a 24-hour day period we have a dummy variable; note that each weekday (Monday, Tuesday, etc.) is allowed to have a potentially different 24-hour profile of intra-day activity. Finally, $\psi_d^{(h)}$ is a dummy variable equal to one if the day of campaign is equal to $d = \{-30, -29, \dots 0, \dots, 29, 30\}$.

Note that in this specification, the lag polynomial of humans supporting campaign Y ($\ln Human_{t-k,d'}^{(Y)}$) starts with $k = 1$ while the lag polynomials for bots start at $k = 0$. This timing means that we allow bots to have a contemporaneous effect on humans and bots do not respond to humans. Consistent with earlier studies using the Jordà approach, we use Newey-West standard errors to account for serial correlation of the error term for $h \geq 1$. We use $K = 24$ for the reported results but our findings are largely unchanged for alternative values of K.

We compute the impulse response to bots supporting campaign X as $\left\{\hat{\alpha}_{X,0}^{(h)}\right\}_{h=0}^{H}$ and the impulse response to bots supporting campaign Y as $\left\{\hat{\beta}_{X,0}^{(h)}\right\}_{h=0}^{H}$. Note that we use logs in specification (1) to transform the volume of tweeting activity (this helps to make the distribution of tweet volume better behaved) so that α_0 and β_0 are elasticities. To convert these elasticities into "multipliers" (that is, a tweet from bot X generates N tweets by humans supporting X), we multiply α by the ratio $\overline{Human^{(X)}}/\overline{Bot^{(X)}} \approx \overline{(Human^{(X)}/Bot^{(X)})}$, that is, the time-series average of the $Human_d^{(X)}/Bot_d^{(X)}$ ratio. Correspondingly, the multiplier from bot Y to human X is the product of β and $\overline{(Human^{(X)}/Bot^{(Y)})}$.

One can think of the Jordà method as constructing a moving average representation of a series: the lag polynomial terms control for initial conditions while $\left\{\hat{\alpha}_{X,0}^{(h)}\right\}_{h=0}^{H}$ and $\left\{\hat{\beta}_{X,0}^{(h)}\right\}_{h=0}^{H}$ describe the behavior of the system in response to a structural, serially uncorrelated shock. Indeed, if we abstract from variation in initial conditions at time t, we effectively regress a variable of interest at time $t + h$ on a shock in a given regime at time t and thus we obtain an average response of the variable of interest h periods after the shock, which is precisely the definition of an impulse response.

As discussed in Jordà (2005) and Auerbach and Gorodnichenko (2012), this approach has several advantages over vector autoregressions (VARs). First, it obviates the need to estimate the

equations for dependent variables other than the variable of interest and thus we can significantly economize on the number of estimated parameters. Second, it does not constrain the shape of the impulse responses. Third, one can easily test joint hypotheses about paths of estimated impulse response. Finally, specification (1) may be straightforwardly extended in various dimensions to allow for a larger set of controls or for more flexible (potentially non-linear) responses. For example, we are interested in comparing the strength of a reaction to human posts supporting campaign Y to the strength of a reaction to bot posts supporting campaign Y. To obtain this comparison, we can estimate

$$\begin{aligned}\ln Human_{t+h,d}^{(X)} = &\sum_{k=0}^{K} \alpha_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(X)} + \sum_{k=0}^{K} \beta_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(Y)} \\ &+ \sum_{k=1}^{K} \gamma_{X,k}^{(h)} \ln Human_{t-k,d'}^{(X)} + \sum_{k=0}^{K} \phi_{X,k}^{(h)} \ln Human_{t-k,d'}^{(Y)} \\ &+ \psi_{X,d}^{(h)} + Seasonal_{X,td}^{(h)} + error_{X,td}^{(h)} \qquad (1')\end{aligned}$$

where now the lag polynomial for $\ln Human_{t-k,d'}^{(Y)}$ starts at $k = 0$ rather than $k = 1$ and one can use $\left\{\hat{\phi}_{X,0}^{(h)}\right\}_{h=0}^{H}$ as the impulse response of humans supporting X to humans supporting Y and corresponding measure multipliers are the product of ϕ and $\overline{(Human^{(X)}/Human^{(Y)})}$. Note that this specification is equivalent to ordering $Human^{(Y)}$ before $Human^{(X)}$ in a VAR.

3.2.2. Baseline results

Figure 7 reports estimated impulse responses for the Brexit sample. Panel A of the figure shows the reaction of humans supporting the leave campaign to messages generated by bots supporting the leave campaign and by bots supporting the remain campaign. The response to "remain" bots is generally small with a weak reaction on impact and a modest, positive multiplier in subsequent periods. In contrast, the contemporaneous reaction of "leave" humans to "leave" bots is strong: the multiplier is close to 2, that is, a new bot post generates two new human posts. However, this elevated tweeting activity of humans is short-lived: after approximately 2-4 hours of the bot post we observe little difference in the response of "leave" humans to "leave" bots and to "remain" bots. These patterns are similar to the behavior of humans in retweeting posts thus providing additional evidence of fast news cycles and/or low information rigidity.

Panel B of the figure plots the responses of "remain" humans to "leave" bots and "remain" bots. Similar to what we observe in Panel A, the reaction of "remain" humans to bots advocating the other side of the campaign is rather mute (the multiplier is close to zero), while the reaction to bots from the same side of the campaign is stronger (the multiplier is about 0.7 which is smaller in absolute terms than the contemporaneous multiplier for "leave" humans in response to "leave" bots). Likewise, the ability of "remain" bots to predict "remain" humans is rather transitory. The asymmetric response of humans to posts consistent/inconsistent with their views suggests that social media can create "echo chambers" fostering amplification of messages within a group of similarly minded people and inhibiting communication of people with different views.

The patterns are similar for the 2016 U.S. Presidential Election (Figure 8). Human supporters of the Trump (Clinton) campaign are more reactive to messages posted by bots supporting the Trump (Clinton) campaign than to messages posted by bots supporting the Clinton (Trump) campaign. In a similar spirit, the reactions are not persistent and most of the response happens within a few hours after a message appears.

[Figures 7 and 8 go here]

In addition to the general trends, the network structure of a user (i.e. networks of followers and the accounts that the user follows) also plays a role in information transmission. For example, a tweet of an influential user, who has a large number of followers, is more likely to be viral. At the same time, a user who has a wide following network is likely to be exposed to more (diverse) information, and thus, facing the question of which information to consume. These, in return, can affect the degree of information spread. While our data do not allow us to examine the network structure in detail[19], we attempt to control for the effect of network structure by accounting for the number of followers in estimating models (1) and (1') and get similar results.[20,21]

[19] In our data, we can only observe the number of followers and following each user has by the time of creating the tweet, but not the full network i.e. who follows whom.

[20] We scale the numbers of tweets created by a user within a time interval by the ratio of the user's number of followers to the maximum number of followers within that time interval.

[21] Appendix D uses standard time-series tools to compute the historical contribution of bot traffic to human users' traffic. Appendix E reports correlations between Twitter traffic and vote outcomes in the U.S and U.K.

3.2.3. News published by traditional media as a confounding factor

One could argue that our baseline results do not necessarily show bots' influence on humans' tweeting activities because we may observe a coincidence of bots' and humans' reactions to news stories published by traditional media outlets. To check the robustness of our findings to this potentially confounding force, we modify model (1) to include a new variable measuring the volume of news generated by conventional media outlets:

$$\begin{aligned}\ln Human_{t+h,d}^{(X)} = &\sum_{k=0}^{K} \alpha_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(X)} + \sum_{k=0}^{K} \beta_{X,k}^{(h)} \ln Bot_{t-k,d'}^{(Y)} \\ &+ \sum_{k=1}^{K} \gamma_{X,k}^{(h)} \ln Human_{t-k,d'}^{(X)} + \sum_{k=1}^{K} \phi_{X,k}^{(h)} \ln Human_{t-k,d'}^{(Y)} \\ &+ \sum_{k=0}^{K} \delta_{X,k}^{(h)} \ln News_{t-k,d'} + \psi_{X,d}^{(h)} + Seasonal_{X,td}^{(h)} + error_{X,td}^{(h)} \qquad (2)\end{aligned}$$

where $\ln News_{t-k,d}$ is the volume of event-related news published during the $t-k$ ten-minute interval on day d' where $d' = d$ if the $t-k$ interval is on the same day with t and $d' = d-1$ if the $t-k$ interval is on the day proceeding t. To be conservative, we use the lag polynomial for $\ln News$ starting at $k = 0$. This variable is constructed from the database of RSS feeds of (1) U.S. media outlets including major national media (e.g., ABC, AP, CNBC, Fox News, New York Times, Wall Street Journal, Washington Post) and local media and (2) 6 U.K. media outlets including BBC, The Guardian, The Independent, Metro, Daily Mail, and The Sun. For the U.S. Election sample, there are 50 media outlets that have news stories related to the election during the investigated time span (30 days before and after the Election day). We define a piece of news as Election-related if its title or summary contains one of the following keywords: Election, Trump, or Clinton. For Brexit sample, a piece of news is defined as Brexit related if its title or summary contains one of the following keywords: Brexit, EU Referendum, Vote Leave, or Vote Remain.

If both bots and humans react to actual news stories and bots happen to react faster than humans, the contemporaneous reaction of humans to the informational shocks generated by bots should be weakened. We find that our baseline results are not sensitive to controlling for the volume of news (Appendix Figures B6 and B7): impulse responses in model (2) are very similar to impulse responses in model (1).

3.2.4. Sentiment

As discussed above, the intensity of human responses may vary with the strength of the sentiment in messages posted by bots. To study this possible heterogeneity in responses, we modify specification (1) as follows:

$$\begin{aligned}\ln Human_{t+h,d}^{(\tau,X)} = &\sum\nolimits_{s\in\{positive,neutral,negative\}}\sum_{k=0}^{K}\alpha_{X,s\to\tau,k}^{(h)}\ln Bot_{t-k,d'}^{(s,X)} \\ &+\sum\nolimits_{s\in\{positive,neutral,negative\}}\sum_{k=0}^{K}\beta_{X,s\to\tau,k}^{(h)}\ln Bot_{t-k,d'}^{(s,Y)} \\ &+\sum\nolimits_{s\in\{positive,neutral,negative\}}\sum_{k=1}^{K}\gamma_{X,s\to\tau,k}^{(h)}\ln Human_{t-k,d'}^{(s,X)} \\ &+\sum\nolimits_{s\in\{positive,neutral,negative\}}\sum_{k=1}^{K}\phi_{X,s\to\tau,k}^{(h)}\ln Human_{t-k,d'}^{(s,Y)} \\ &+\psi_{X,\tau,d}^{(h)}+Seasonal_{X,\tau,td}^{(h)}+error_{X,\tau,td}^{(h)} \qquad (3)\end{aligned}$$

where τ and s index sentiment (tonality) from positive to neutral to negative. One can now interpret slopes as the strength of the response that may vary by tonality of messages posted by humans and bots. For example, $\alpha_{X,s\to\tau,k}^{(h)}$ measures the elasticity of the h-horizon response of humans' tweets supporting campaign X and expressing sentiment τ to bot posts supporting campaign X with sentiment s. Then we use appropriate ratios of the regressand to a regressor to convert the estimated elasticities into multipliers. Figure 9 and Figure 10 plot the estimated impulse responses (measured as multipliers) for the U.K. and U.S. samples respectively. By and large, we observe results similar to the baseline results: humans supporting a given side of a campaign tend to react stronger to posts generated by bots supporting the same side and the sentiment of human responses mimics the sentiment of bot posts.

[Figures 9 and 10 go here]

3.2.5. Humans vs. bots

Our analysis so far has presented evidence consistent with the view that the Twitter activity of bots can affect the Twitter activity of humans who share beliefs advocated by bots. However, a bot

appears to have a weak predictive power for humans' activity who have beliefs opposite to what is advocated by the bot. Would humans be more effective in reaching across the aisle?

To answer this question, we use specification (1') to compare response multipliers for humans who support a given campaign side to posts by humans and bots from the other camp. Panel A of Figure 11 shows response multipliers of "leave" humans in the Brexit sample to tweets posted by "remain" humans and "remain" bots. We observe that, if anything, bots appear to have larger multipliers than humans. Likewise, "remain" humans appear to have larger multipliers in response to "leave" bots than to "leave" humans (Panel B). The results are similar for the U.S. sample (Figure 12): bots appear to be as effective (if not more effective) as humans in moving humans with opposite views.

[Figures 11 and 12 go here]

To further examine the ability of bots to "persuade" humans to switch sides, we perform the following analysis. In each sample, we obtain a sub-sample of tweets of "active" humans who created at least 20 original tweets.[22] Three tweets' elements are used for our analysis: (1) the timing of tweets, (2) whether it is an original tweet or a retweet from bots, (2) which side is supported in the tweet. A human account is considered "switching sides" if we observe that account doing a sequence like this: at time t-2, an original tweet is created and is pro-Trump (pro-leave) tweet; at time t-1, the user retweets a pro-Clinton (pro-remain) tweet which was generated by a bot; and at time t, an original tweet is created and is pro-Clinton (pro-remain) tweet. Alternatively, a user doing the following sequence is also considered as "switching sides": at time t-2, an original tweet is created and is pro-Clinton (pro-remain) tweet; at time t-1, the user retweets a pro-Trump (pro-leave) tweet which was created by a bot; and at time t, an original tweet is created and is pro-Trump (pro-leave) tweet. We find that in the Brexit sample about 10% of human accounts have ever switched sides. This statistic is around 6% for the U.S. Election sample. Thus, these results are consistent with the notion that humans can switch sides after being exposed to bots with contrary views but this behavior is exhibited by a small share of humans.

These results suggest that while human touch and personal connections may be important, in the world of social media bots and other "strangers" can play an equally important role in bringing

[22] 5,325 human accounts in the Brexit sample are selected and 14,124 human accounts in the U.S. Election sample are selected.

together or distancing people with different beliefs. Although bots could sway opinions of only a small share of humans, one should bear in mind that bot traffic is considerably cheaper than traffic generated by humans (Forelle et al., 2015). Hence, one may anticipate ever greater use of bots in political campaigns as well as the increasing effectiveness of bots in changing humans' opinions. At the same time, one might also expect various attempts of humans to shield themselves from bots.

4. Concluding remarks

Social media are a powerful tool for spreading news and information. However, social media might also propagate misinformation and fake news, especially during high-impact events. It is necessary to understand how information is diffused and acquired in social networks as it might affect individuals' decision-making in real life. Furthermore, the rise of bots (automated agents) in social media potentially creates greater risks of manipulation as humans may not detect bots and thus could be affected and possibly deceived by bots.

This study explores the diffusion of information on Twitter during two high-impact political events in the U.K. (2016 E.U. Referendum, "Brexit") and the U.S. (2016 Presidential Election). Specifically, we empirically examine how information flows during these two events and how individuals' actions might be influenced by different types of agents. We have two key results. First, information about the Brexit Referendum and the U.S. Election is disseminated quickly on Twitter. During these two highly covered campaigns, reaction to new messages is largely complete within 1-2 hours which is consistent with fast news cycles and/or low information rigidity in social media.

Second, we find that individuals are more active in interacting with similar-minded Twitter users. That is, pro-leave users react faster and stronger to the messages created by other pro-leave users. We document that the extent to which bots' tweeting activities can predict humans' tweeting activities likely depends on whether bots' information is consistent with humans' preferences. For example, a message by a pro-leave bot generates a response of pro-leave humans and approximately no response of pro-remain humans. Furthermore, bots' messages with a given sentiment largely generate human messages with the same sentiment. These results lend support to the "echo chambers" view that Twitter and other social media create networks for individuals sharing similar political beliefs so that they tend to interact with others from the same communities

and thus their beliefs are reinforced. By contrast, information from outsiders is more likely to be ignored. Consequently, ideological polarization in social media like Twitter is likely amplified rather than attenuated, which makes reaching consensus on important public issues more difficult.

Since Twitter and other platforms of social media may create a sense of public consensus or support, social media could indeed affect public opinions in new ways. Specifically, social bots could spread and amplify (mis)information thus influencing what humans think about a given issue and likely reinforcing humans' beliefs. Not surprisingly, bots were used during the two campaigns we study to energize voters. These two campaigns and subsequent debates about the role of bots in shaping the campaigns highlight the importance of having policies/mechanisms to prevent abuse of bots in social media platforms (see, e.g., UK Parliament, 2018; European Commission, 2019). In fact, since 2018, social media platforms like Twitter or Facebook have started internal audits to remove fake/bot accounts. However, despite this ongoing effort, the presence of social media bots is still an (if not greater) issue since (1) automated accounts are "cheap" to produce and (2) bot accounts are increasingly sophisticated and thus, are more difficult to be detected (e.g., Deb et al., 2019; Luceri et al., 2019).

Because we rely on observational data, we cannot determine causal effects beyond reasonable doubt. Subsequent work should aim to establish the causal effects of bots on humans' activities as well as the underlying mechanisms, of which results will help social media platforms have better control over the use of bot accounts. For instance, one can gain better understanding of humans' awareness of the presence of bot accounts and their potential influence through large-scale surveys. Conscious efforts/collaboration from social media platforms are also required to implement randomized control trials for other research designs (e.g., difference-in-differences, regression discontinuity). In doing so, one will be able to disentangle different channels through which bots can affect humans (e.g., direct channels when humans follow and retweet from bots versus indirect channels when humans are exposed to bots through "mutual friends").

Obviously, regulating information flows is an extremely delicate business in a democratic society characterized by diverse views and tolerance for this diversity. However, cherishing diversity does not mean that one should allow the spread of lies and manipulation that work to prevent the public from making well-informed decisions. Where one should draw the line (e.g., disclose ultimate owners of accounts, improve media literacy, or introduce "code of practice" for social networks) is a central question for society.

References

Acemoglu, D., Hassan, T.A. and Tahoun, A., 2018. The power of the street: Evidence from Egypt's Arab Spring. *Review of Financial Studies*, *31*(1), pp.1-42.

Falck, O., Gold, R. and Heblich, S., 2014. E-lections: Voting Behavior and the Internet. *American Economic Review*, *104*(7), pp.2238-65.

Caetano, J.A., Lima, H.S., Santos, M.F. and Marques-Neto, H.T., 2018. Using Sentiment Analysis to Define Twitter Political Users' Classes and Their Homophily During the 2016 American Presidential Election. *Journal of Internet Services and Applications*, *9*(1), pp.1-15.

Adamic, L.A. and Glance, N., 2005, August. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, pp. 36-43, ACM.

Auerbach, A.J. and Gorodnichenko, Y., 2012. Fiscal multipliers in recession and expansion. In *Fiscal Policy after the Financial Crisis*, pp. 63-98, University of Chicago press.

Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A. and Bonneau, R., 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, *26*(10), pp.1531-1542.

Bastos, M.T. and Mercea, D., 2017. The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C. and Vohs, K.D., 2001. Bad is stronger than good. *Review of General Psychology*, *5*(4), p.323.

Benoit, K., 2017. The Brexit Debate through Social Media: Deliberative Discussion, or Deliberate Dysfunction?. [Online] London School of Economics and Political Science. Available at: http://www.lse.ac.uk/Events/2017/01/20170125t1830vHKT/The-Brexit-Debate-through-Social-Media [Accessed 29 April 2018].

Bermingham, A. and Smeaton, A., 2011. On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, pp. 2-10.

Bessi, A. and Ferrara, E., 2016. Social bots distort the 2016 US Presidential election online discussion. *First Monday*, *21*(11).

Bialik, K. and Matsa, K., 2017. Key trends in social and digital news media. [Online] Pew Research Center. Available at: http://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/ [Accessed 29 April 2018].

Burnap, P., Gibson, R., Sloan, L., Southern, R. and Williams, M., 2016. 140 characters to victory?: Using Twitter to predict the UK 2015 General Election. *Electoral Studies*, *41*, pp.230-233.

Chen, Z. and Lurie, N.H., 2013. Temporal contiguity and negativity bias in the impact of online word of mouth. *Journal of Marketing Research*, *50*(4), pp.463-476.

Chu, Z., Gianvecchio, S., Wang, H. and Jajodia, S., 2010. Who is tweeting on Twitter: human, bot, or cyborg?. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pp. 21-30, ACM.

Chu, Z., Gianvecchio, S., Wang, H. and Jajodia, S., 2012. Detecting automation of twitter accounts: Are you a human, bot, or cyborg?. *IEEE Transactions on Dependable and Secure Computing*, *9*(6), pp.811-824.

Coibion, O. and Gorodnichenko, Y., 2012. What can survey forecasts tell us about information rigidities?. *Journal of Political Economy*, *120*(1), pp.116-159.

Coibion, O., Gorodnichenko, Y., Kueng, L. and Silvia, J., 2017. Innocent Bystanders? Monetary Policy and Inequality. *Journal of Monetary Economics*, *88*, pp.70-89.

Cook, D.M., Waugh, B., Abdipanah, M., Hashemi, O. and Rahman, S.A., 2014. Twitter deception and influence: Issues of identity, slacktivism, and puppetry. *Journal of Information Warfare*, *13*(1), pp.58-IV.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J. and Neuberger, C., 2013. An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. *Information, Communication & Society*, *16*(5), pp.795-825.

Davis, Clayton A., Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer, 2016. "BotOrNot: A System to Evaluate Social Bots," Manuscript. Available at https://arxiv.org/pdf/1602.00975.pdf.

Deb, A., Luceri, L., Badaway, A. and Ferrara, E., 2019, May. Perils and Challenges of Social Media and Election Manipulation Analysis: The 2018 US Midterms. In Companion Proceedings of The 2019 World Wide Web Conference (pp. 237-247). ACM.

DellaVigna, S. and Kaplan, E., 2007. The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, *122*(3), pp.1187-1234.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M. and Zhuravskaya, E., 2014. Cross-border media and nationalism: Evidence from Serbian radio in Croatia. *American Economic Journal: Applied Economics*, *6*(3), pp.103-132.

Druckman, J.N., Levendusky, M.S. and McLain, A., 2018. No need to watch: How the effects of partisan media can spread via interpersonal discussions. *American Journal of Political Science*, *62*(1), pp.99-112.

Enikolopov, R., Makarin, A. and Petrova, M., 2019. Social media and protest participation: Evidence from Russia. *Econometrica, Forthcoming*.

Enikolopov, R., Petrova, M. and Sonin, K., 2018. Social media and corruption. *American Economic Journal: Applied Economics*, *10*(1), pp.150-74.

European Commission, 2019. Tackling online disinformation - Digital Single Market - European Commission. [online] Available at: https://ec.europa.eu/digital-single-market/en/tackling-online-disinformation [Accessed 11 Dec. 2019].

Ferrara, E., Varol, O., Davis, C., Menczer, F. and Flammini, A., 2016. The rise of social bots. *Communications of the ACM*, *59*(7), pp.96-104.

Forelle, M.C., Howard, P.N., Monroy-Hernandez, A. and Savage, S., 2015. Political Bots and the Manipulation of Public Opinion in Venezuela. *SSRN Electronic Journal*.

Franzese, Robert J., Jr. 2002. "Electoral and Partisan Cycles in Economic Policies and Outcomes," *Annual Review of Political Science* 5(1), 369-421.

Garrett, R.K., 2009a. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, *14*(2), pp.265-285.

Garrett, R.K., 2009b. Politically motivated reinforcement seeking: Reframing the selective exposure debate. *Journal of Communication*, *59*(4), pp.676-699.

Gentzkow, M. and Shapiro, J.M., 2011. Ideological segregation online and offline. *The Quarterly Journal of Economics*, *126*(4), pp.1799-1839.

Gentzkow, M., 2006. Television and voter turnout. *The Quarterly Journal of Economics*, *121*(3), pp.931-972.

Gerber, A.S., Karlan, D. and Bergan, D., 2009. Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, *1*(2), pp.35-52.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J. and Moreno, Y., 2014. Assessing the bias in samples of large online networks. *Social Networks*, *38*, pp.16-27.

Gottfried, J. and Shearer, E., 2016. News Use Across Social Media Platforms 2016. [Online] Pew Research Center's Journalism Project. Available at:

http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/ [Accessed 29 April 2018].

Gruzd, A. and Roy, J., 2014. Investigating political polarization on Twitter: A Canadian perspective. *Policy & Internet*, 6(1), pp.28-45.

Gruzd, A., Wellman, B. and Takhteyev, Y., 2011. Imagining Twitter as an imagined community. *American Behavioral Scientist*, *55*(10), pp.1294-1318.

Halberstam, Y. and Knight, B., 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, *143*, pp.73-88.

Hargittai, E., Gallo, J. and Kane, M., 2008. Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, *134*(1-2), pp.67-86.

Haustein, S. and Costas, R., 2014. Determining Twitter audiences: Geolocation and number of followers. *ALM*, *4*, p.6.

Haustein, S., Bowman, T.D., Holmberg, K., Tsou, A., Sugimoto, C.R. and Larivière, V., 2016. Tweets as impact indicators: Examining the implications of automated "bot" accounts on Twitter. *Journal of the Association for Information Science and Technology*, *67*(1), 232-238.

Heath, C., 1996. Do people prefer to pass along good or bad news? Valence and relevance of news as predictors of transmission propensity. *Organizational Behavior and Human Decision Processes*, *68*(2), pp.79-94.

Hong, S. and Kim, S.H., 2016. Political polarization on twitter: Implications for the use of social media in digital governments. *Government Information Quarterly*, *33*(4), pp.777-782.

Howard, P.N. and Kollanyi, B., 2016. Bots, #StrongerIn, and #Brexit: Computational Propaganda during the U.K.-EU Referendum. Working Paper 2016. *Computational Propaganda Project: Working Paper Series*.

Howard, P.N., Kollanyi, B. and Woolley, S., 2016. Bots and Automation over Twitter during the US Election. *Computational Propaganda Project: Working Paper Series*.

Howard, P.N., Woolley, S. and Calo, R., 2018. Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, *15*(2), pp.81-93.

Jordà, Ò., 2005. Estimation and inference of impulse responses by local projections. *American Economic Review*, *95*(1), pp.161-182.

Kim, J. and Yoo, J., 2012, December. Role of sentiment in message propagation: Reply vs. retweet behavior in political communication. In *2012 International Conference on Social Informatics*, pp. 131-136, IEEE.

Kwak, H., Lee, C., Park, H. and Moon, S., 2010. What is Twitter, a social network or a news media?. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 591-600, ACM.

Lee, K., Caverlee, J. and Webb, S., 2010. Uncovering social spammers: social honeypots+ machine learning. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 435-442, ACM.

Lelkes, Y., Sood, G. and Iyengar, S., 2015. The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, *61*(1), pp.5-20.

Levendusky, M.S., 2013. Why do partisan media polarize viewers?. *American Journal of Political Science*, *57*(3), pp.611-623.

Lopez, J.C.A.D., Collignon-Delmar, S., Benoit, K. and Matsuo, A., 2017. Predicting the Brexit vote by tracking and classifying public opinion using Twitter data. *Statistics, Politics and Policy*, *8*(1), pp.85-104.

Loria, S., 2018. TextBlob Documentation. [Online] Available at: https://media.readthedocs.org/pdf/textblob/dev/textblob.pdf [Accessed 29 April 2018].

Luceri, L., Deb, A., Giordano, S. and Ferrara, E., 2019. Evolution of bot and human behavior during elections. First Monday, 24(9).

Meffert, M.F., Chung, S., Joiner, A.J., Waks, L. and Garst, J., 2006. The effects of negativity and motivated information processing during a political campaign. *Journal of Communication*, *56*(1), pp.27-51.

Morstatter, F., Pfeffer, J., Liu, H. and Carley, K.M., 2013, July. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *ICWSM*.

Mullainathan, S. and Shleifer, A., 2005. The Market for News. *American Economic Review*, *95*(4), pp.1031-1053.

Müller, K. and Schwarz, C., 2018. Fanning the Flames of Hate: Social Media and Hate Crime (No. 373). Competitive Advantage in the Global Economy (CAGE).

New York Times, 2018. Battling Fake Accounts, Twitter to Slash Millions of Followers. [Online] Available at: https://www.nytimes.com/2018/07/11/technology/twitter-fake-followers.html [Accessed 11 Jun. 2019].

O'Connor, B., Balasubramanyan, R., Routledge, B.R. and Smith, N.A., 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media*.

Peisakhin, L. and Rozenas, A., 2018. Electoral effects of biased media: Russian television in Ukraine. *American Journal of Political Science*, *62*(3), pp.535-550.

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

Sivanesh, S., Kavin, K. and Hassan, A.A., 2013, August. Frustrate Twitter from automation: How far a user can be trusted?. In *Human Computer Interactions (ICHCI), 2013 International Conference,* pp. 1-5, IEEE.

Steinert-Threlkeld, Z.C., Mocanu, D., Vespignani, A. and Fowler, J., 2015. Online social networks and offline protest. *EPJ Data Science*, *4*(1), p.19.

Stempel, G.H., Hargrove, T. and Bernt, J.P., 2000. Relation of growth of use of the Internet to changes in media use from 1995 to 1999. *Journalism & Mass Communication Quarterly*, *77*(1), pp.71-79.

Stieglitz, S. and Dang-Xuan, L., 2013. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), pp.217-248.

Stukal, D., Sanovich, S., Bonneau, R. and Tucker, J.A., 2017. Detecting bots on Russian political Twitter. *Big Data*, *5*(4), pp.310-324.

Takhteyev, Y., Gruzd, A. and Wellman, B., 2012. Geography of Twitter networks. *Social Networks*, *34*(1), pp.73-81.

Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M., 2011. Election forecasts with Twitter: How 140 characters reflect the political landscape. *Social Science Computer Review*, *29*(4), pp.402-418.

UK Parliament, 2018. House of Lords - The politics of polling - Select Committee on Political Polling and Digital Media. [online] Available at: https://publications.parliament.uk/pa/ld201719/ldselect/ldppdm/106/10609.htm [Accessed 11 Dec. 2019].

Veale, T., Valitutti, A. and Li, G., 2015. Twitter: The best of bot worlds for automated wit. In International Conference on Distributed, Ambient, and Pervasive Interactions, pp. 689-699. Springer, Cham.

Velez, Y.R. and Newman, B.J., 2019. Tuning In, Not Turning Out: Evaluating the Impact of Ethnic Television on Political Participation. *American Journal of Political Science*.

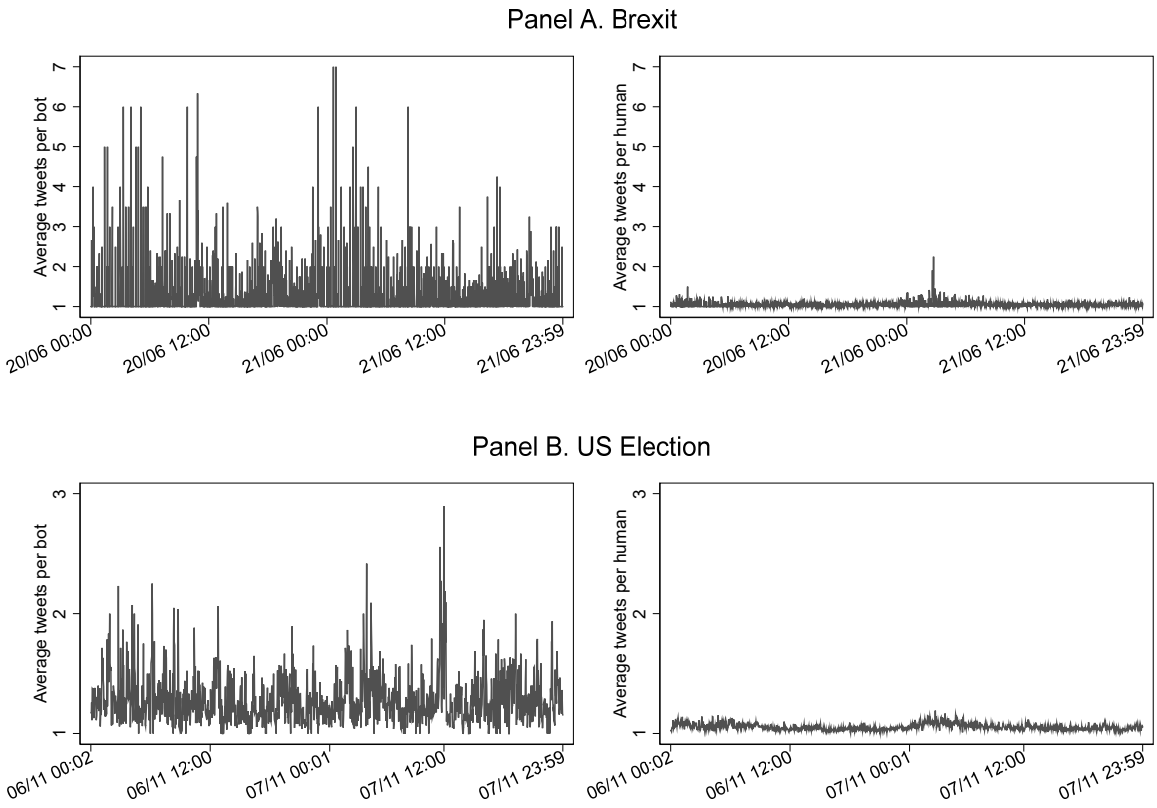
Vosoughi, S., Roy, D. and Aral, S., 2018. The spread of true and false news online. Science, *359*(6380), pp.1146-1151.

Wojcik, S., 2018. 5 things to know about bots on Twitter. [Online] Pew Research Center. Available at: http://www.pewresearch.org/fact-tank/2018/04/09/5-things-to-know-about-bots-on-twitter/ [Accessed 31 Jul. 2018].

Woolley, S.C. and Howard, P., 2017. Computational propaganda worldwide: Executive summary.

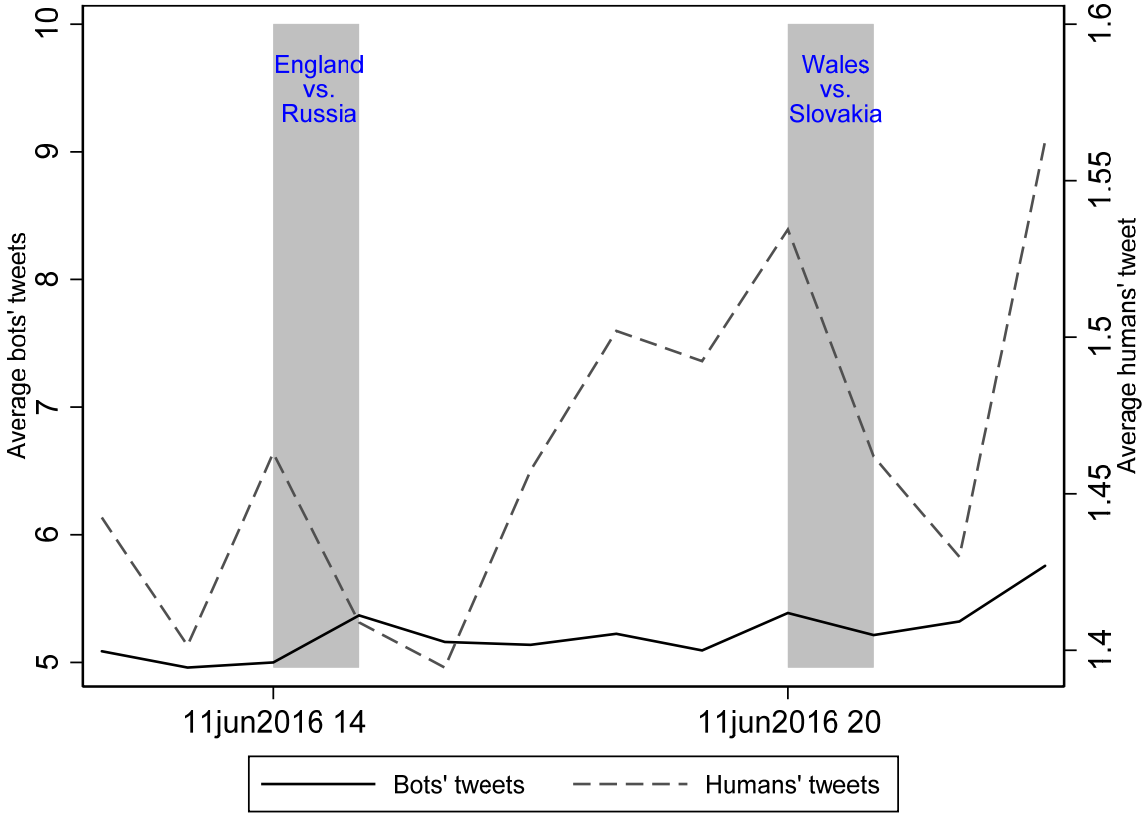
Yoo, E., Rand, W., Eftekhar, M. and Rabinovich, E., 2016. Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management*, *45*, pp.123-133.

Figure 1. Average tweets generated by humans and bots per minute



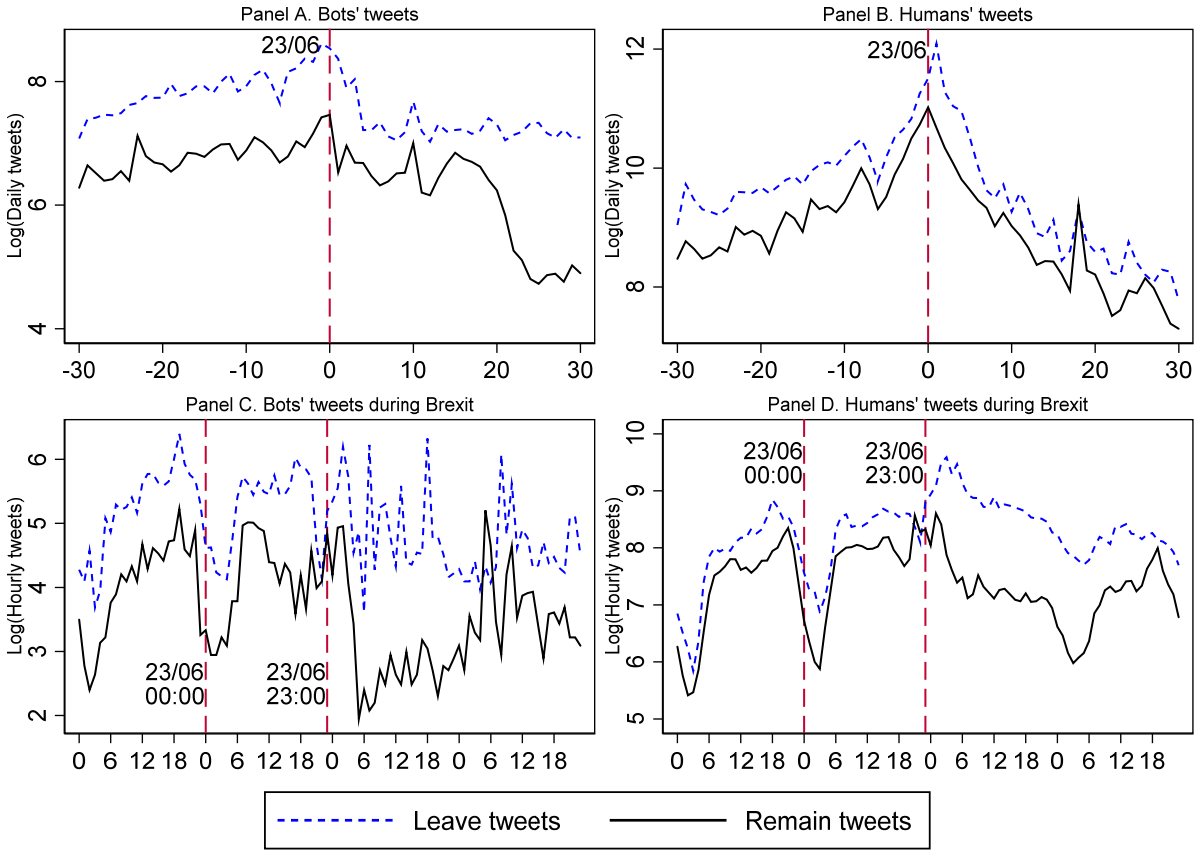
Notes: This figure shows the average number of tweets per account generated within 1-minute interval over 2 days prior to the Brexit Referendum day (Panel A) and the 2016 U.S. Presidential Election day (Panel B). The left panels show the average tweets generated by bots and the right panels show the average tweets created by humans.

Figure 2. Bots' and humans' Brexit tweets during football matches



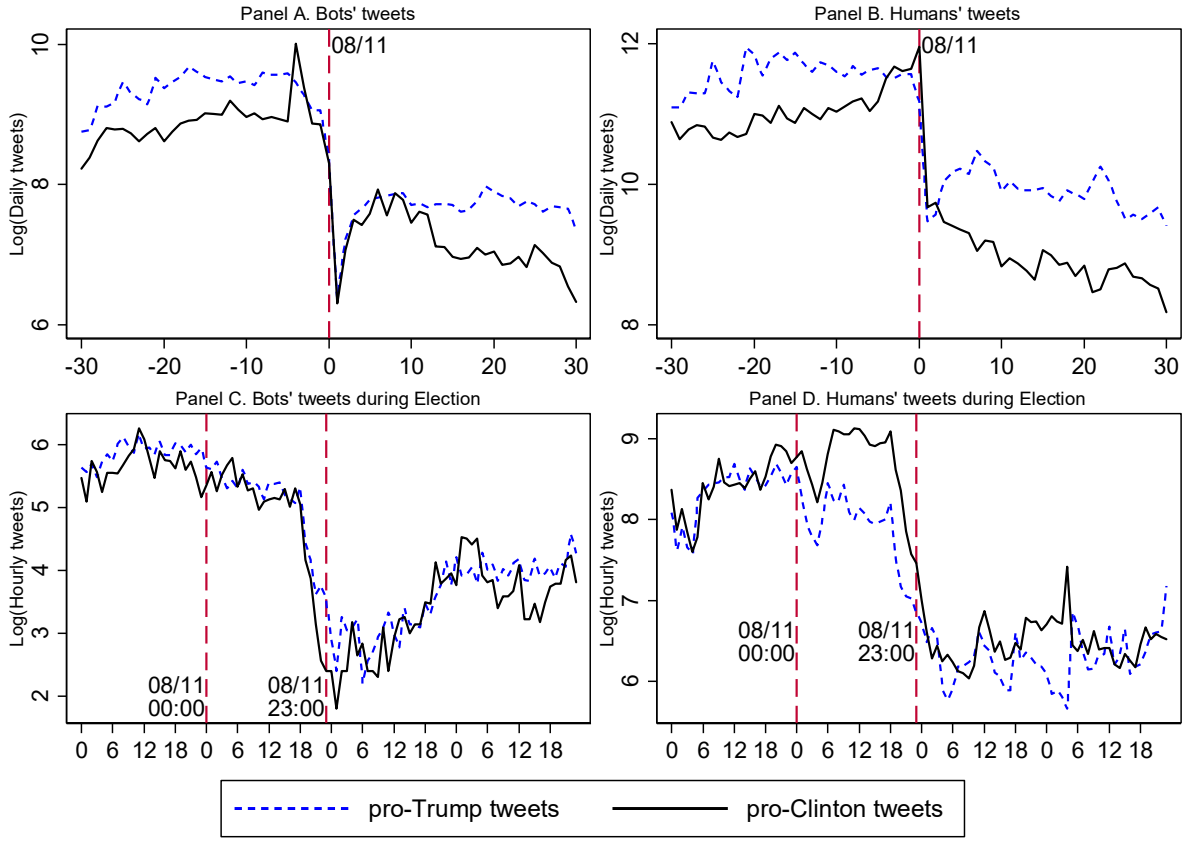
Notes: This figure shows the average number of hourly tweets generated by humans and bots in Brexit sample during 12:00 and 23:00 on 11th June 2016. The solid line shows the average tweets generated by bots while the dashed line shows the average tweets created by humans. The grey shaded areas represent the matchs' duration. The first match is between England and Russia which took place at 14:00. The second match is between Wales and Slovakia which took place at 20:00.

Figure 3. Dynamics of tweets generated by humans and bots: U.K. Brexit



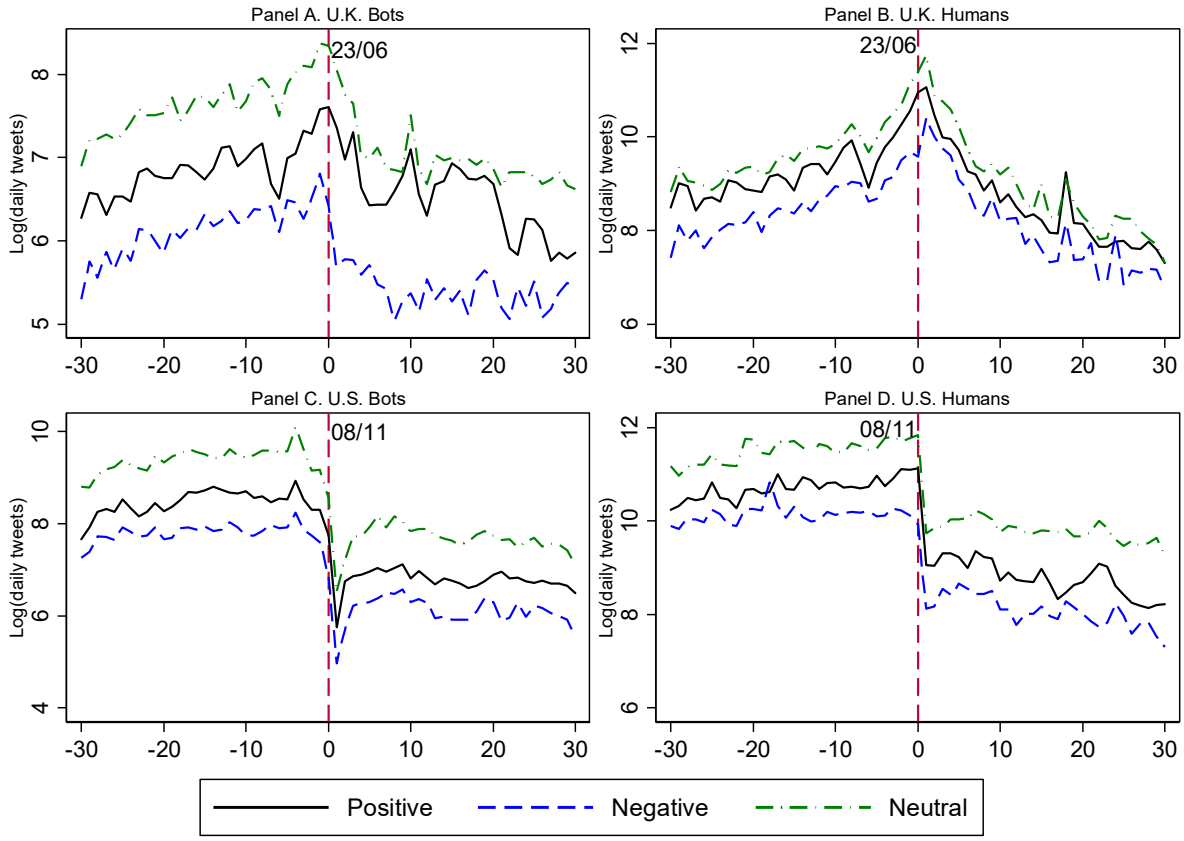
Notes: This figure shows the dynamics of tweets created by bots (Panels A and C) and humans (Panels B and D) for the Brexit sample. Time (horizontal axis) in Panels A and B shows 30 days before and after the Referendum day while time (horizontal axis) in Panels C and D presents hours of days around the event. The dashed blue line and the solid black line show the volumes of pro-leave and pro-remain tweets, respectively.

Figure 4. Dynamics of tweets generated by humans and bots: 2016 U.S. Presidential Election



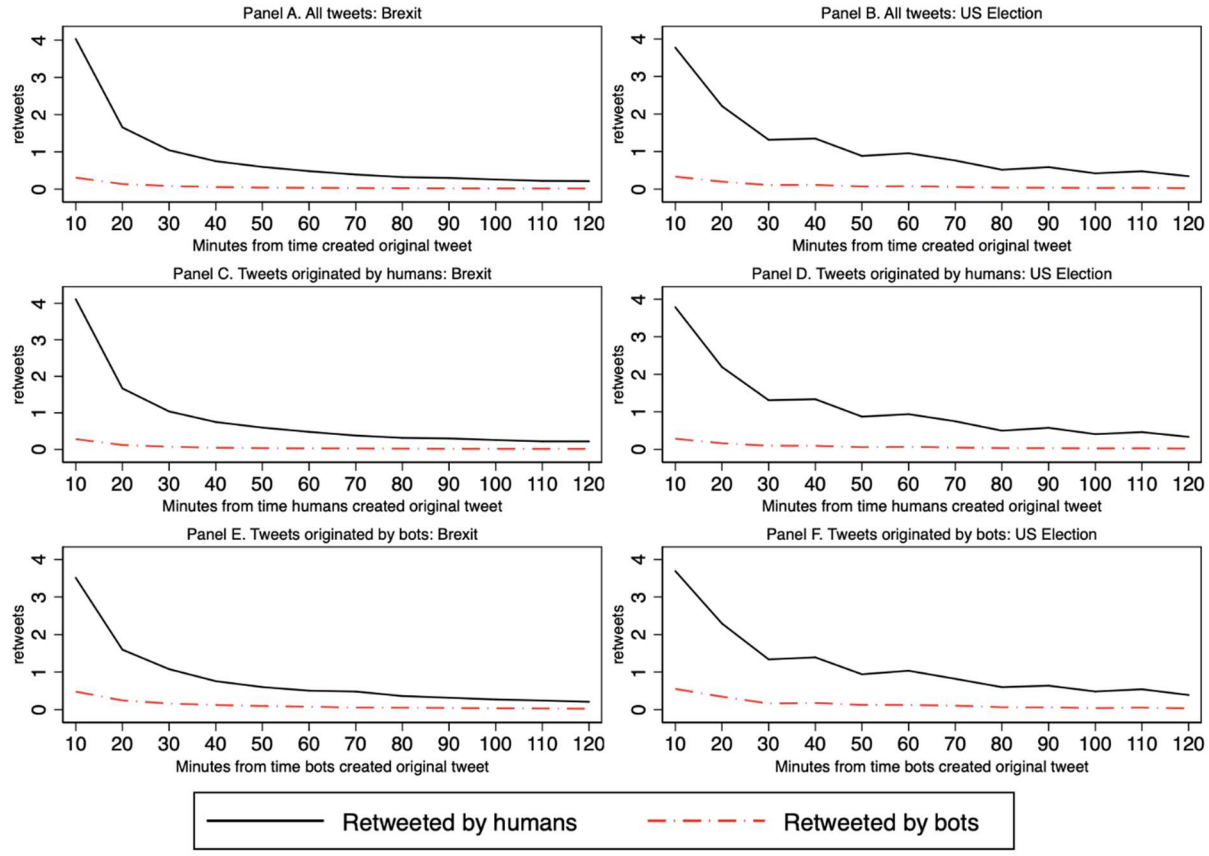
Notes: This figure shows the dynamics of tweets created by bots (Panels A and C) and humans (Panels B and D) for the 2016 U.S. Presidential Election sample. Time (horizontal axis) in Panels A and B shows 30 days before and after the Election Day while time (horizontal axis) in Panels C and D presents hours of days around the event. The dashed blue line and the solid black line show the volumes of pro-Trump and pro-Clinton tweets, respectively.

Figure 5. Sentiment



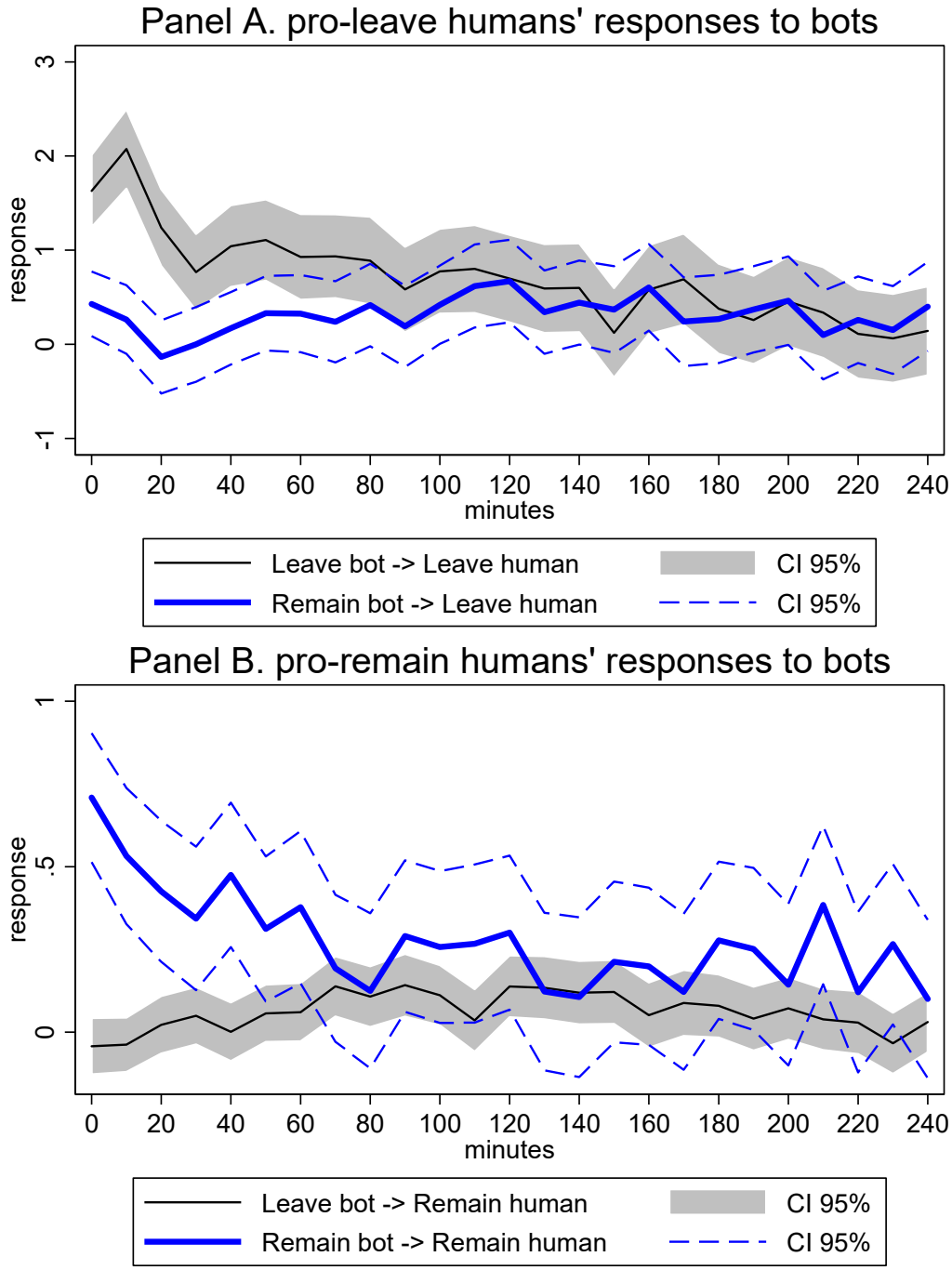
Notes: This figure shows the dynamics of tweets with different sentiment created by bots and humans for the samples of Brexit (Panels A and B) and the 2016 U.S. Presidential Election (Panels C and D). Time (horizontal axis) shows 30 days before and after the event day. The dashed blue line, the solid black line, and the dashed-dotted green line show the volumes of tweets with negative, positive, and neutral sentiment, respectively.

Figure 6. Retweeting activity by humans and bots



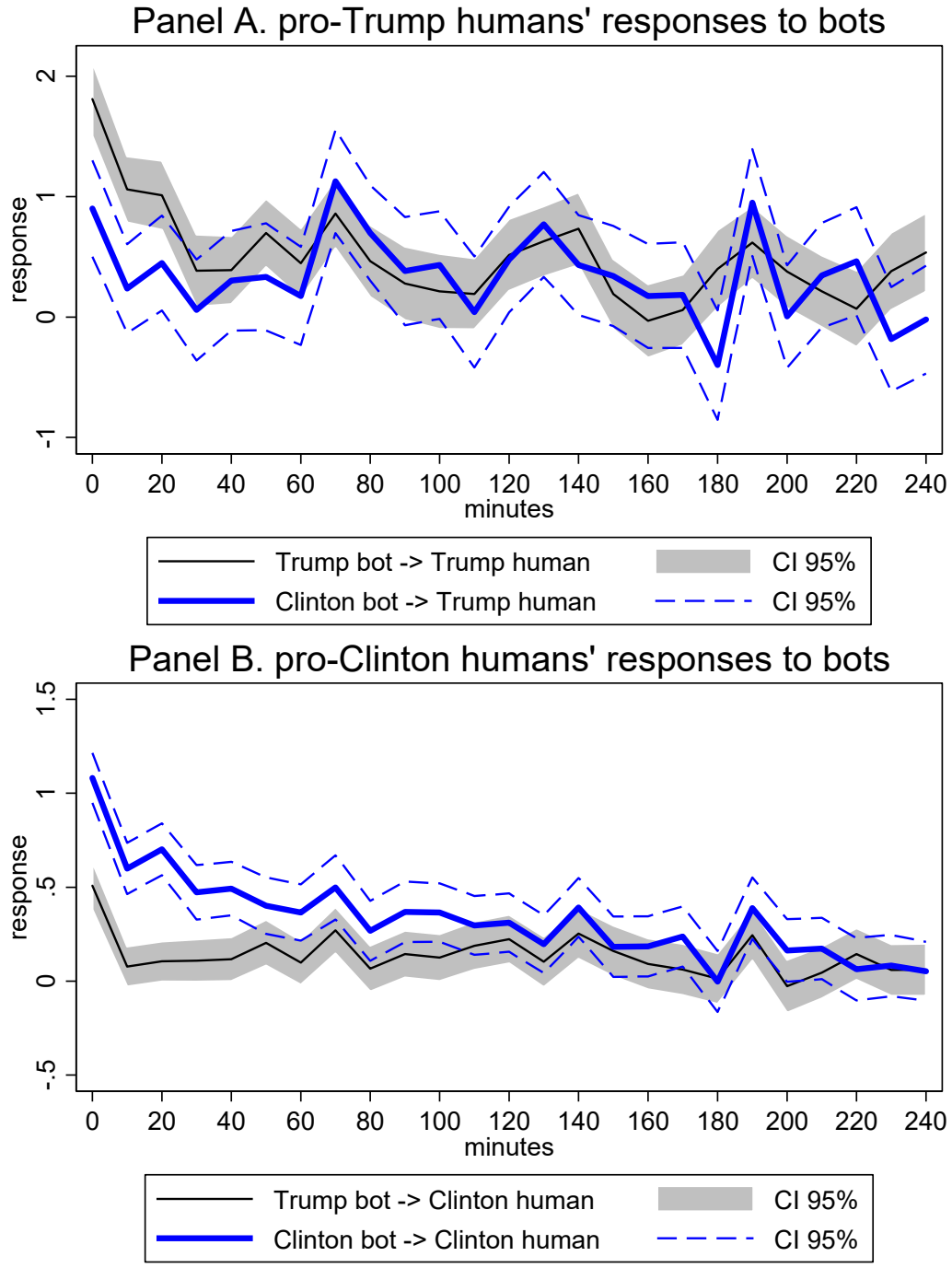
Notes: This figure shows the average number of retweets made over a 10-120 minutes period after each original tweet was created for the Brexit sample (Panels A, C, E) and the 2016 U.S. Presidential Election sample (Panels B, D, F). Panels A and B show the average retweets of all tweets for each sample. Panels C and D show the average retweets of the tweets originated by humans. Panels E and F show the average retweets of the tweets originated by bots. The solid black line refers to the retweeting activities of humans while the dashed red line refers to the retweeting activities of bots. Time (horizontal axis) is in 10-minute intervals.

Figure 7. U.K. (baseline)



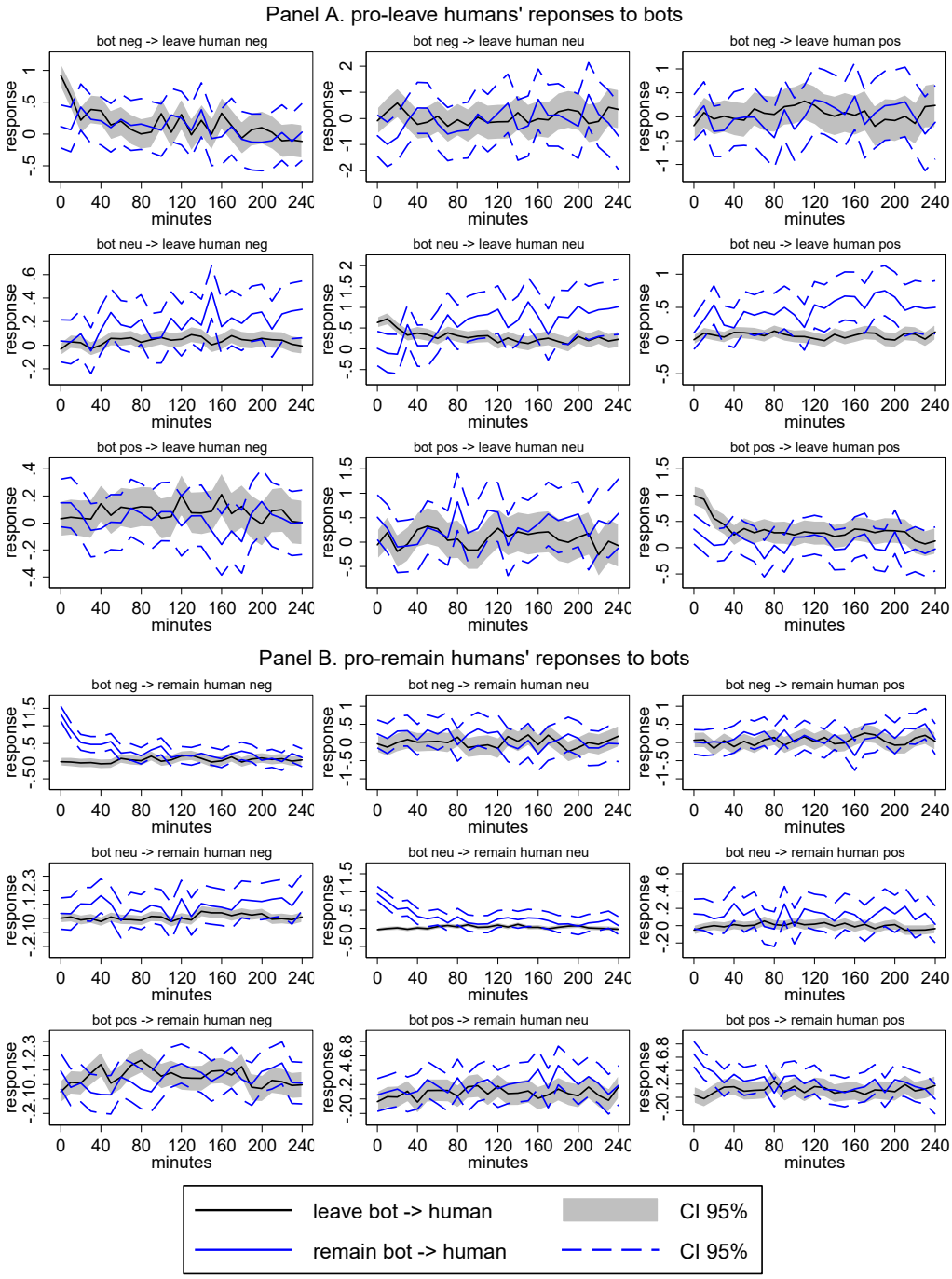
Notes: This figure reports estimated impulse responses of humans to tweets created by bots for the Brexit sample. Panels A and B show the reactions of humans supporting the leave campaign and the remain campaign, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions to pro-leave bots while the thick blue line refers to the reactions to pro-remain bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-leave bots and pro-remain bots, respectively.

Figure 8. U.S. (baseline)



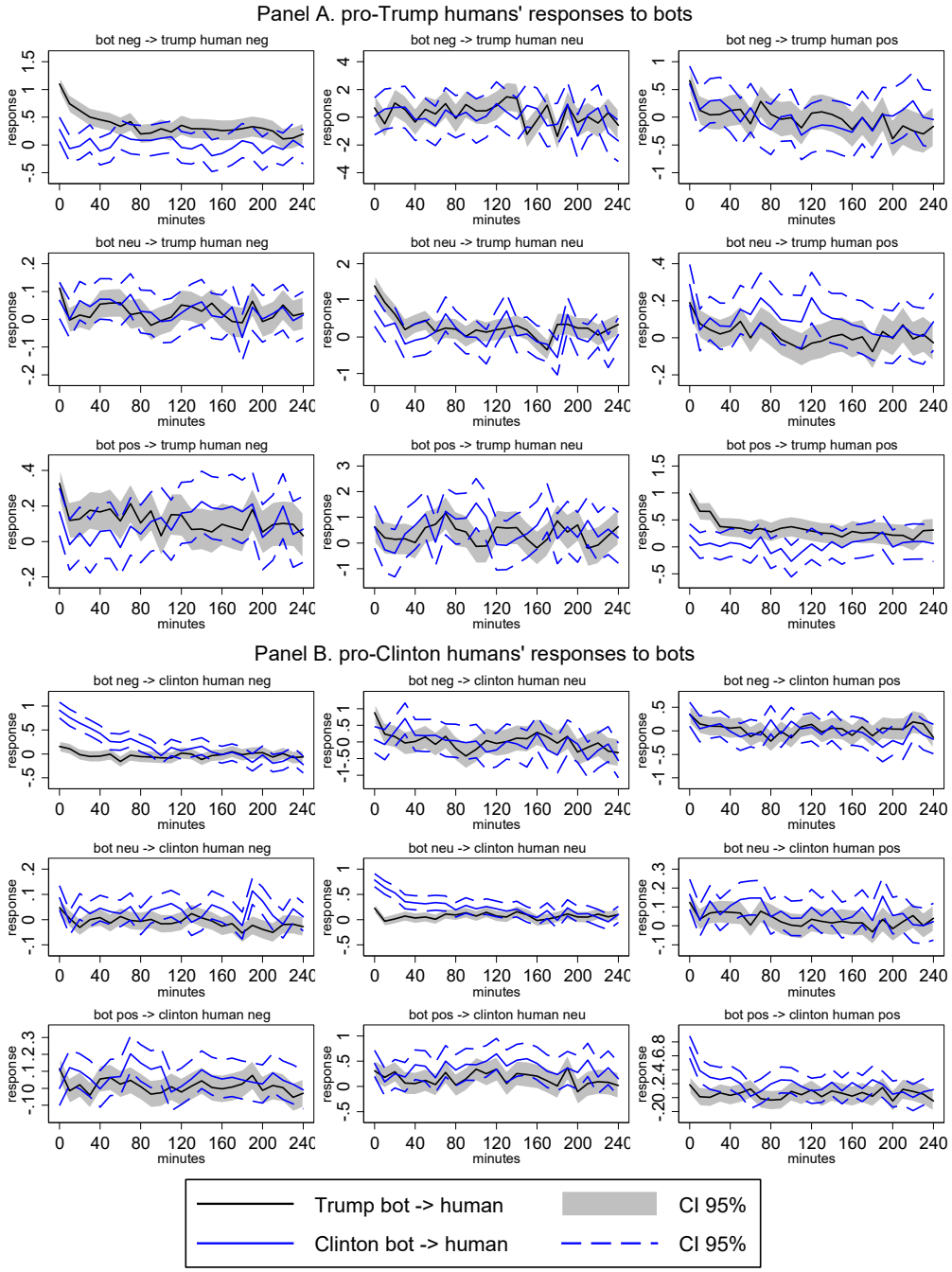
Notes: This figure reports estimate impulse responses of humans to tweets created by bots for the 2016 U.S. Presidential Election sample. Panels A and B show the reactions of humans supporting Trump and Clinton, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions to pro-Trump bots while the thick blue line refers to the reactions to pro-Clinton bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-Trump bots and pro-Clinton bots, respectively.

Figure 9. U.K. (by sentiment)



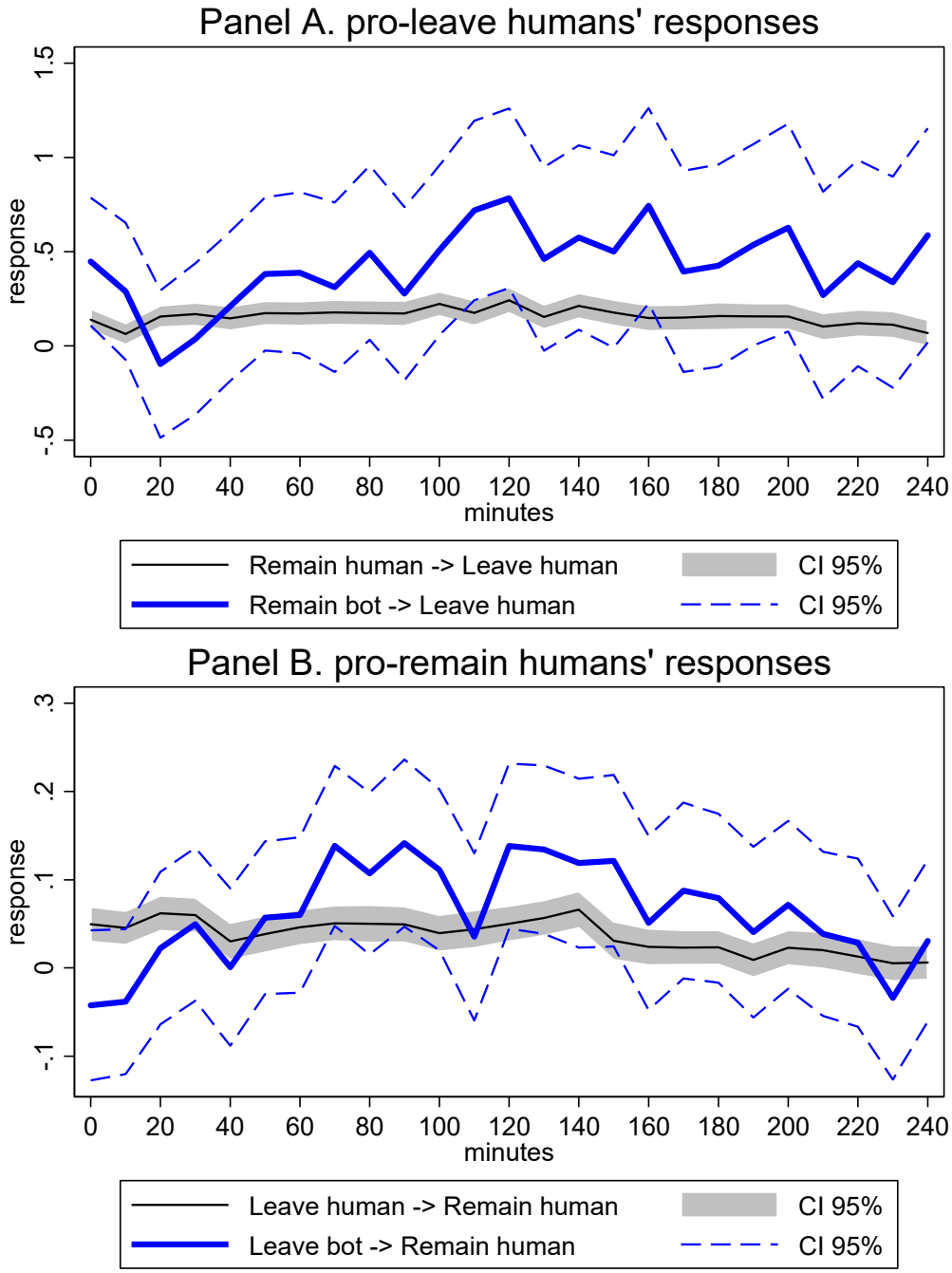
Notes: This figure reports the estimated impulse responses (measured as multipliers) of humans to bots' tweets with different sentiment for the Brexit sample. Panels A and B show reactions of humans supporting the leave campaign and the remain campaign, respectively. Time (horizontal axis) is in 10-minute intervals. The solid black line refers to the reactions to pro-leave bots while the solid blue line refers to the reactions to pro-remain bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-leave bots and pro-remain bots, respectively. *neg*, *neu*, and *pos* stand for negative, neutral, and positive sentiment respectively.

Figure 10. U.S. (by sentiment)



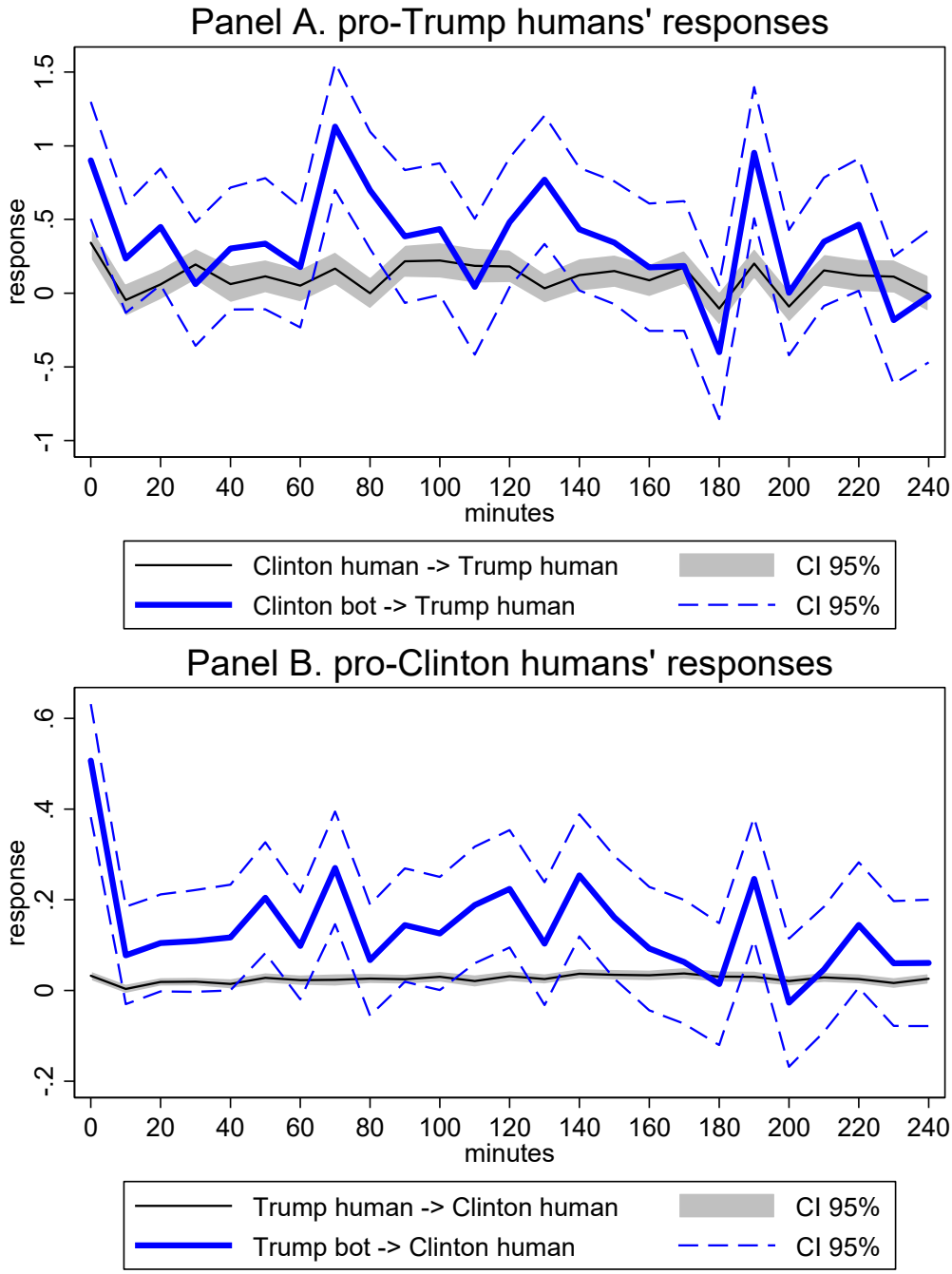
Notes: This figure reports the estimated impulse responses (measured as multipliers) of humans to bots' tweets with different sentiment for the 2016 U.S. Presidential Election sample. Panels A and B show reactions of humans supporting Trump and Clinton, respectively. Time (horizontal axis) is in 10-minute intervals. The solid black line refers to the reactions to pro-Trump bots while the solid blue line refers to the reactions to pro-Clinton bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-Trump bots and pro-Clinton bots, respectively. *neg*, *neu*, and *pos* stand for negative, neutral, and positive sentiment respectively.

Figure 11. U.K.: humans vs bots in effectiveness to generate tweets of the other side



Notes: This figure reports the estimated impulse responses (measured as multipliers) of humans to tweets supporting the opposite campaign for the Brexit sample. Panels A and B show reactions of humans supporting the leave campaign and the remain campaign, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions of humans to humans while the thick blue line refers to the reactions of humans to bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to tweets created by humans and bots, respectively.

Figure 12. U.S.: humans vs bots in effectiveness to generate tweets of the other side



Notes: This figure reports the estimated impulse responses (measured as multipliers) of humans to tweets supporting the opposite campaign for the 2016 U.S. Presidential Election sample. Panels A and B show reactions of humans supporting Trump and Clinton, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions of humans to humans while the thick blue line refers to the reactions of humans to bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to tweets created by humans and bots, respectively.

Table 1. Descriptive statistics by bot and human accounts

	Brexit		U.S. Election	
	Bots	Humans	Bots	Humans
	(1)	(2)	(3)	(4)
No. of accounts	850	586,879	1,754	871,432
Age	1,206	1,580	944	1,467
Followers	4,646	3,130	1,071	2,316
Statuses	34,581	14,903	18,784	12,627
Daily volume of original tweets	15.5	1.0	10.4	1.6
Daily volume of retweets	15.2	0.6	11.4	0.5

Notes: This table presents summary statistics for bot and human accounts in the Brexit and U.S. Election sample. *Age* is the average account age (in days) as of the end of each sample (23 July 2016 for the Brexit sample and 08 December 2016 for the U.S. Election sample). *Followers* and *Statuses* are the average number of followers and statuses, respectively. *Daily volume of original tweets* indicates the average number of the daily original tweets created by an account. *Daily volume of retweets* indicates the average number of the daily retweets created by an account.

ONLINE SUPPLEMENTAL APPENDIX

Appendix A.

Table A1. Hashtags/keywords used to define tweets

Pro-leave tweets	leave, voteleave, leaveeu, takecontrol, betteroffout, voteout, beleave, brexitthemovie, euistheproblem, brexitbustour, strongerout
Pro-remain tweets	strongerin, remain, voteremain, votein, bremain, votestay, intogether, labourinforbritain, greenerin
Pro-Trump tweets	americanfirst, imwithyou, lawandorder, makeamericagreatagain, neverClinton, draintheswamp, teamtrump, crookedClinton, trumpwon, lockherup, trumptrain, fortrump, clintonfoundation, trump2016, sendhertojail, Clintonforprison, votetrump
Pro-Clinton tweets	imwithher, lovetrumpshate, nevertrump, clintonkaine, ohhillyes, strongertogether, dirtydonald, Clintonwon, votedems, trumpeduptrickledown, whyiwantClinton, proClinton, Clintonsarmy, tntweeters, uniteblue, womenstoptrumpparty, dumptrump, donaldtrumpthemovie, trumpsajoke, trumpsacrifices, makedonalddrumpfagain, forClinton, lasttimetrumppaidtax, boycotttrump, Clintonsupporter

Notes: This table shows the hashtags/keywords used to define supporting side during the 2016 E.U. Referendum and the 2016 U.S. Presidential Election.

Table A2. Process of defining original tweets and their retweets

	Step 1: Create retweet indicator	Step 2: Information about the user from whom the tweet was retweeted	
Text	RT	RT_from	Tweet type
RT @jameswilson: RT @abcnews: After this brexit campaign europe cannot remain the same	1	jameswilson	Retweet
After this brexit campaign europe cannot remain the same	0		Original tweet
RT @abcnews: After this brexit campaign europe cannot remain the same	1	abcnews	Retweet

Notes: In this table, we show an example of how we define retweets in our data.

Appendix A3. Example of sentiment analysis using Textblob

To be clear about the sentiment analysis discussed in Section 2.4, let's consider this sentence: "It is interesting and funny" which has two adjectives ("funny" and "interesting"). The word "funny" has 4 entries in TextBlob's pre-classified lexicon with polarity scores of 0.5, 0, 0.5, and 0, making the word's average polarity score 0.25. The word "interesting" has only 1 entry in the lexicon with the polarity score of 0.5. Thus, the polarity score of the sentence is 0.375 ($=\frac{0.25+0.5}{2}$). The similar process is applied in analyzing sentiment of this sentence "It is interesting and very funny" with the only difference is that "very" is a modifier of which only its intensity score (1.3) is taken into account. In this case, the phrase "very funny" has the polarity score of 0.325 ($= 0.25 \times 1.3$). Hence, the polarity score of the whole sentence is 0.4125 (= (0.325+0.5)/2).

Table A3. Examples of tweets with polarity scores from TextBlob

Text	Polarity score
RT @Nigel_Farage: I don't believe these official figures and I'm sure the real numbers are much higher. We must Leave EU and control our border https://t.co/3pr.	0.32
Brexit may break Britain's Tory party	0
Is there a prize for being the most ironic idiot on the planet?	-0.033
I don't believe those official figures and I'm sure the real numbers are much higher. We must Leave EU and control our border	0.317
Brexit frees the Tories to destroy workers' right	-0.25
If we allow the 18 Albanians who arrived on our shores illegally to stay then UK faces a migrant crisis this summer	-0.5
I'm voting leave because nobody owns allegiance to an unelected elite	0
To All Trump Voters You are my brothers and sisters. Thank you for giving America more reasons for hope and joy this Thanksgiving	0.65
Donald Trump doesn't like poor people	-0.4
NBCSandiego and also JAIL THE CLINTON MAFIA lol lyinmedia	0.35
Trump's showing himself to be corrupt AF and he's only getting started folks	-0.25
PresidentElect Trump Releases Thanksgiving Message Focused on Unity VIDEO gatewaypundit	0

Notes: In this table, we show some examples of polarity score obtained from TextBlob.

Appendix A4. Examples of bot accounts

In this section, we will provide more examples of bot accounts and their characteristics as well as samples of tweets created by these accounts. Note that the tweets provided here are after-cleaning tweets e.g. after removing URL links etc.

Example 1: Account **julieca92128716** was created on May 6, 2016. By Oct 7, 2016, this account had created 47,531 tweets (on average more than 200 tweets per day!). However, the account has stopped tweeting since 2018 (the last tweet was on April 10, 2018). A sample of this account's tweets is provided below.

Date	Time	Text
8-Oct-16	07:10:44	RT @daystotrump 32 days to trump counting down to the election in november makeamericagreatagain
8-Oct-16	07:26:02	RT @terrymendozer podestaemails wikileaks guess whats trending off twitter hint not the trump tape womenfortrump
8-Oct-16	07:26:02	RT @doyawannabfree so thee only choice for usa is blacksfortrump fwd
8-Oct-16	07:30:24	RT @marypatriott hlary clinton my dream is open borderspodestaemails neverhillary maga tcot
8-Oct-16	08:05:21	RT @trumpworld trump trump2016trumptrain
8-Oct-16	08:07:32	RT @waynedupreeshow i am voting realdonaldtrump over hillaryclinton in november debates2016 votetrump neverhillary
8-Oct-16	13:48:19	RT @mitchellvii so the nevertrumps really think trump should step down over some private macho locker room talk 11 years ago ridiculou
8-Oct-16	13:50:30	RT @loudobbs maga trumppence16 mikepence realdonaldtrump americafirst trumptrain dobbs
8-Oct-16	14:53:52	RT @carrieksada ill just leave this here for you beta males and imwithher folks maga gatewaypundit
8-Oct-16	15:02:36	RT @alisonspalding2 basketofdeplorablesbucketoflosersneverhillarysurrenderhillarytoastpoorhillarylesserintellects
8-Oct-16	15:11:20	RT @jrcheneyjohn wikileaks hillary will impose gun control via executive order bigleaguetruth neverhillary maga trump2016
8-Oct-16	22:54:27	RT @fredfortrump realdonaldtrump crookedclintons crookedhillary maga trump clintonsforprison
8-Oct-16	22:56:38	RT @proudamerican wakeupamerica neverforget neverhillary votetrump2016 followthemoney makeamericagreatagain
9-Oct-16	07:30:00	RT @mitchellvii these nevertrump assholes are willing to turn america over to the evil hillary because trump talked sex hell w them al
9-Oct-16	08:07:09	RT @dixiefortrump thank goodness the clintonmafia respects women neverhillary
9-Oct-16	08:22:26	RT @americatrump neverhillary
9-Oct-16	08:24:37	RT @veteran4trump trump is the only candidate saying we have to take care of our own people first hes for america first trump2016
9-Oct-16	08:24:37	RT @veteran4trump the saddes day for law enforcement will be if hillary were to become president of the united states neverhillary
9-Oct-16	21:13:34	RT @trumppenceusa1 lockherup debatenight makeamericagreatagain

Example 2: Account **38_28_38** was created on December 20, 2010. By October 7, 2016, this account had created 184,536 tweets (on average 87 tweets per day!). The account has been suspended because it violates the Twitter rules. Below is a sample of tweets created by this account.

Date	Time	Text
7-Oct-16	18:15:14	nevertrumptrumps basketofdeplorables base are racists amp sexists they like him more now rosieperezbklyn cnn
7-Oct-16	18:26:09	nevertrump129fans640twts blakeaspen43 troll spinning lies for trump im shocked patchelio spindr cnn
7-Oct-16	18:28:20	nevertrumpsexual deviant trump is a hypocrite who accuse others of his sins blakeaspen43 spindr patchelio cnn
7-Oct-16	19:16:24	nevertrumpgovpencein wont commentmikepence ashamed in front of his daughter matthewnussbaum politico
7-Oct-16	19:22:57	nevertrumplol trumps fake female hispanic surrogate blow off his latest misogyny sad ajdelgado13 gkoukla
8-Oct-16	01:25:35	RT @jimgeraghty why are you dumb nevertrump holdouts refusing to jump on the trump train as it careens off the bridge towards the deep
8-Oct-16	02:00:32	nevertrumpthe only way for the republican party to salvage itself is to shoot trump into space in a rocket ship h
8-Oct-16	02:26:45	RT @382838 nevertrumpthe only way for the republican party to salvage itself is to shoot trump into space in a rocket ship h
8-Oct-16	10:03:19	nevertrumpturns out trumps historically bad week was better than this one lol sad
8-Oct-16	18:36:41	RT @382838 nevertrumpthis is crazy trumps stealing ppls money
8-Oct-16	20:39:01	RT @382838 imwithherthe clintons marriage is none of our business hillary worked thru the bad times amp respected her marri
8-Oct-16	20:39:01	RT @382838 imwithheri respect hillary for taking the hard road to work on her marriage amp make it a success goldengateblond
8-Oct-16	22:19:30	RT @382838 imwithherthe clintons marriage is none of our business hillary worked thru the bad times amp respected her marri
8-Oct-16	22:19:30	RT @382838 imwithheri respect hillary for taking the hard road to work on her marriage amp make it a success goldengateblond
8-Oct-16	23:03:12	imwithheryour phony trumptrolls scandals amp putin hacks dont impress me deplorable dumbass sad alexwy18
8-Oct-16	23:09:45	RT @382838 imwithherthe clintons marriage is none of our business hillary worked thru the bad times amp respected her marri
9-Oct-16	00:06:33	RT @382838 imwithherthe clintons marriage is none of our business hillary worked thru the bad times amp respected her marri
9-Oct-16	00:17:28	trumptrolls nevertrumpnow ill help your dumb deplorable racist ass out of my mts lt liso32 jamestoupin
9-Oct-16	00:21:50	RT @382838 imwithherthe clintons marriage is none of our business hillary worked thru the bad times amp respected her marri
9-Oct-16	09:32:20	RT @382838 imwithhera feminist who championed womens amp childrens rights for over 40yrs hillary cjhennig realronhoward

Example 3: Account **2053pam** was created on August 24, 2010 and as of May 23, 2016, it had created more than 93,800 tweets (on average, it tweeted about 45 tweets per day!). This account has now been suspended. Some examples of its tweets are as follows.

Date	Time	Text

24-May-16	09:31:05	RT @55massey good shout katie ha ha
24-May-16	09:38:41	RT @labourleave lexit the movie nearly halfway funded v tight schedules help get us there brexit lexit
24-May-16	09:39:09	RT @peterghandley holiday cost is a business decision thats why they go up during school holidays nothing to do with in or out eu
24-May-16	09:41:44	RT @dailyexpress furious tories warn of noconfidence vote in david cameron over made up brexit claims
24-May-16	09:44:42	RT @strongerln your tuesdayterror brexit risk updateif we leave the eu dentists will be free to pull out any teeth they like amp with
24-May-16	13:29:20	RT @charlescrawford brexit gt part of a process towards a measured and fair new basis for europe as a whole
24-May-16	13:41:06	RT @interimspace voteleavemedia so leave have to make the case versus best efforts of ourampeu civil service foreign amp bigcorp interest
24-May-16	13:50:00	dorsetrebel ukip nigelfarage he is one of the few earning his salary and will lose his eu salary for british people when we brexit
24-May-16	13:53:28	RT @robertbranson75 who loses the most from brexit try goldman sachs wsj
24-May-16	13:55:37	RT @petenorth303 there are consequences for squandering democracy voteleave
24-May-16	14:09:52	RT @staranderton brexit cheerleaderinchief nigel farage set to roll into dudley
24-May-16	14:24:13	RT @chrishanseneu bedtime for the british people uncle david makes sure you will have horrible nightmares brexit leaveeu
24-May-16	14:24:59	RT @beleaveuk it was only months ago that camerons government said that the uk would be fine after brexit how times change eh beleave
24-May-16	14:52:19	RT @fight4uk brexit go leaveeu
24-May-16	14:52:34	RT @colconnaughton eu referendum these economists say the eu is evil voteleave
24-May-16	15:05:38	RT @bernardtrolling ideal fuck the eu
24-May-16	15:08:57	RT @hotspur007 far right being anyone that doesnt agree with their marxist agenda get out of this fascist dictatorship brexit
24-May-16	15:11:17	RT @fishingforleave help reclaim our right to fish as an island nationleaveeu brexit
24-May-16	15:12:24	RT @racheljoyce ordinary ppl will definitely benefit from brexit so voteleave
24-May-16	15:13:00	RT @kthopkins given we lack leadership over brexit my revised british bill of human rights for your perusal and acceptance
24-May-16	15:18:16	RT @martenscentre brexit or bremain its all about love actually
24-May-16	15:21:34	RT @dcreesy hey davidcameron if the eu is doing so well whats with the shrinking eutime to brexit and trade freely
24-May-16	15:22:09	RT @expressandstar heres nigelfarage on the brexitbustour campaign in dudleywatch the full video here
24-May-16	15:24:13	RT @britslovepolls eu referendum poll resultsremain 23leave 77via blp onlinebrexit
24-May-16	15:27:23	belgium loving being part of the eu 60000 protestors in brussels happy days strongerin brexit
24-May-16	15:36:42	RT @greycurmudgeon remain or leave the european union what do you think leave
24-May-16	15:55:20	this is paris loving being part of the eu strongerin brexit
24-May-16	15:57:43	RT @2tweetaboutit no wonder they want us to remain germany fears brexit because they would lose 35billion
24-May-16	16:05:23	RT @ahmedumbongo if you are more concerned about the cost of your flight to ibiza than your right to democracy then you are part of the pr
24-May-16	16:17:35	RT @tomswarbrick1 stevehiltonx it would be antidemocratic for cameron to go in face of brexit euref lbc
24-May-16	16:19:09	RT @kthopkins if you didnt get your first choice of primary school place voteleave 250000 came to the uk from europe last year brexit

24-May-16	16:23:15	RT @nanette0803 camerons exguru says voters have been treated like simpletons over brexit like
24-May-16	16:23:28	RT @maturefinancier farage is for the people the establishment hate this popularity
24-May-16	16:24:20	RT @m0nk33y dcclaxton 2053pam and as we both know dodgydave is partial to a holiday or three brexit voteleave
24-May-16	16:31:54	RT @gjon777 the government has been murdered for its report on the economy after brexit even nicola sturgeon poo pood it
24-May-16	16:34:44	RT @sgifforduk george osbornes postbrexit economy is a fantasy land
24-May-16	16:35:21	RT @mrtcharris only 230 i just lost a bet i thought he would have gone for above 300
24-May-16	16:47:00	RT @muppetv2 and the bar gets loweris there no position of trust strongerin will not breakvoteleave and brexit june 23rd

Example 4: Account **gopolympics1** was created on June 20, 2016. 10 days after being created (by June 30, 2016), it had already created more than 14,700 tweets. This account has been also suspended by Twitter. Below is a sample of its tweets.

Date	Time	Text
01-Jul-16	00:12:24	brexit michael gove betrays boris johnson says he isnt up to the job of pm
01-Jul-16	00:12:38	brexit michael gove betrays boris johnson says he isnt up to the job of pm
01-Jul-16	00:12:45	brexit michael gove betrays boris johnson says he isnt up to the job of pm
01-Jul-16	00:12:47	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:43:08	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:43:38	brexit michael gove betrays boris johnson says he isnt up to the job of pm
01-Jul-16	00:43:43	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:06	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:07	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:09	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:11	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:13	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:14	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:15	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:17	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:18	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	00:59:19	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	01:00:05	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	01:13:20	brexit boris johnson betrayed by michael gove theresa may now favourite to become pm
01-Jul-16	01:13:36	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	01:14:12	uk in shock as brexit champion boris johnson betrayed bows out of leadership bid
01-Jul-16	01:14:20	brexit boris johnson betrayed by michael gove theresa may now favourite to become pm
01-Jul-16	01:30:06	brexit michael gove betrays boris johnson says he isnt up to the job of pm
01-Jul-16	01:30:08	brexit michael gove betrays boris johnson says he isnt up to the job of pm

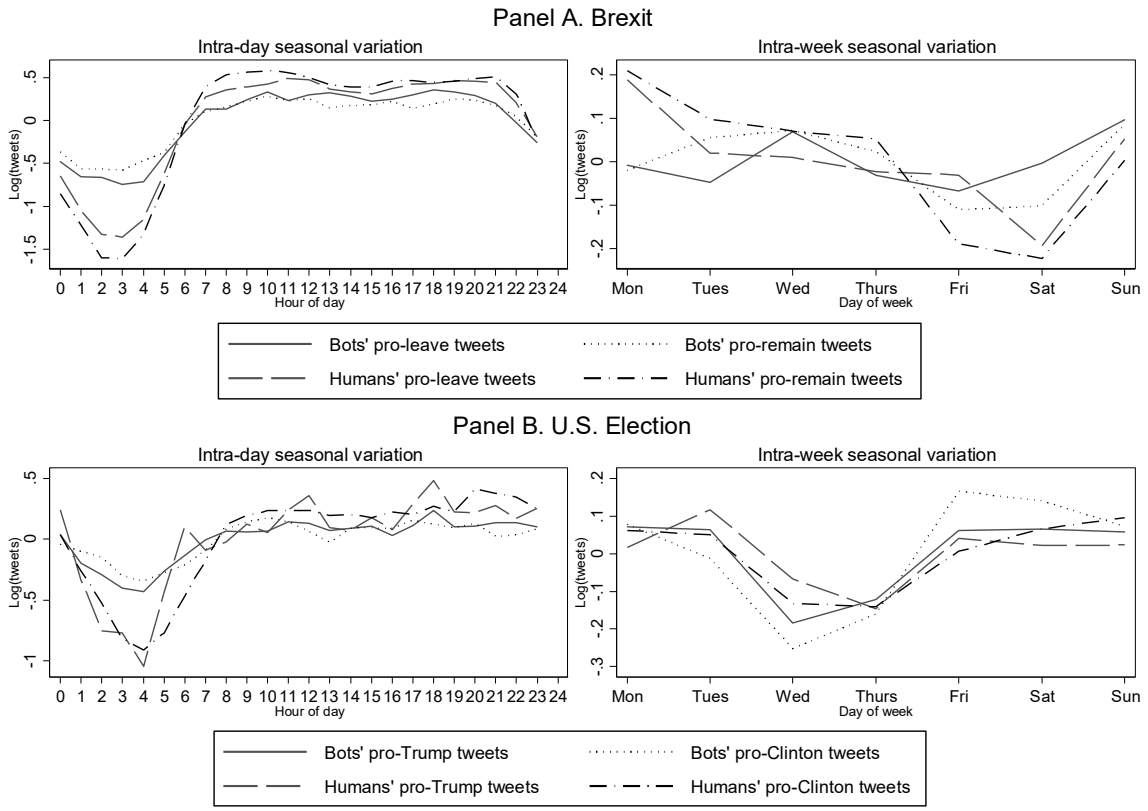
Table A4. Key statistics for estimation samples

	Brexit sample (24 May to 23 July 2016)	U.S. Election sample (09 Oct to 08 Dec 2016)
Campaign endorsement		
No. of pro-leave tweets	1,536,115	
No. of pro-remain tweets	739,536	
No. of pro-Trump tweets		4,353,896
No. of pro-Clinton tweets		2,571,208
Tweet sentiment		
No. of positive tweets	728,619	1,800,557
No. of negative tweets	361,847	985,045
No. of neutral tweets	1,185,185	4,139,502
Tweet classification		
No. of original tweets	845,926	1,911,423
No. of retweets	1,429,725	5,013,681
No. of bots' tweets	178,601	705,262

Notes: This table presents summary statistics for some key variables in our analysis.

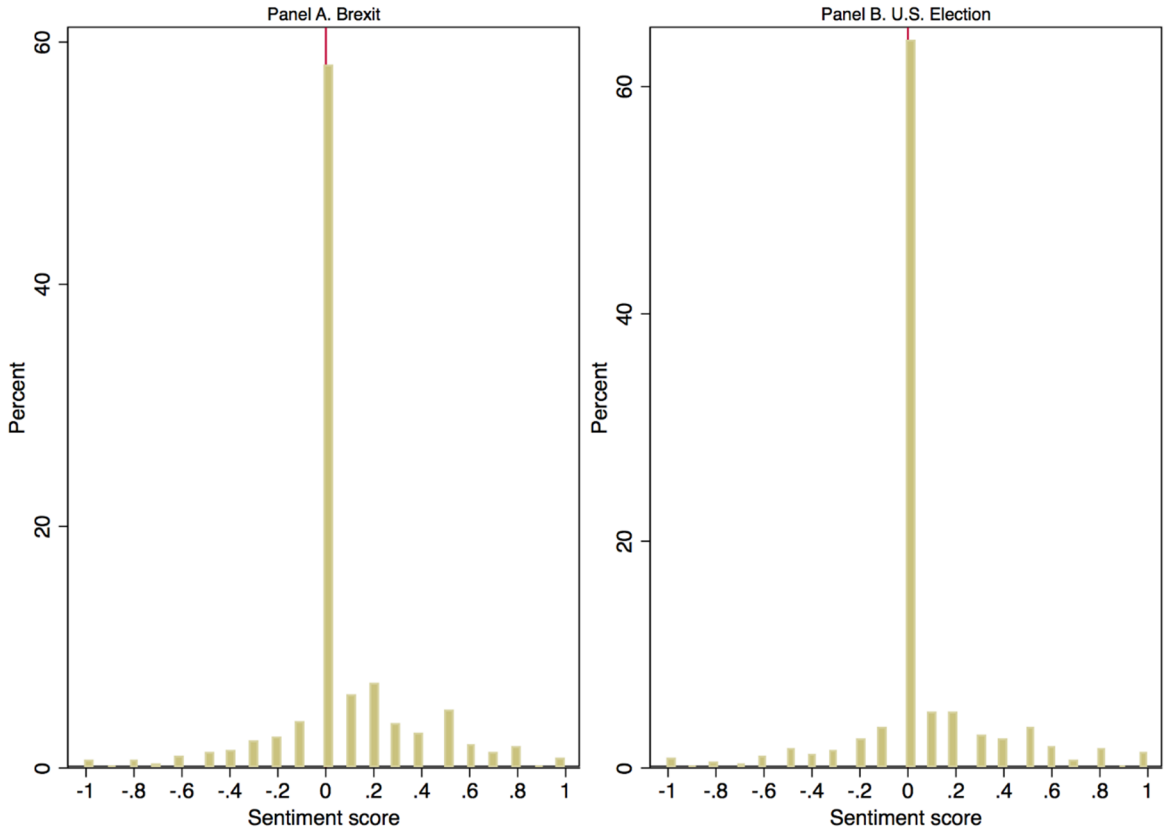
Appendix B. Figures

Figure B1. Seasonality



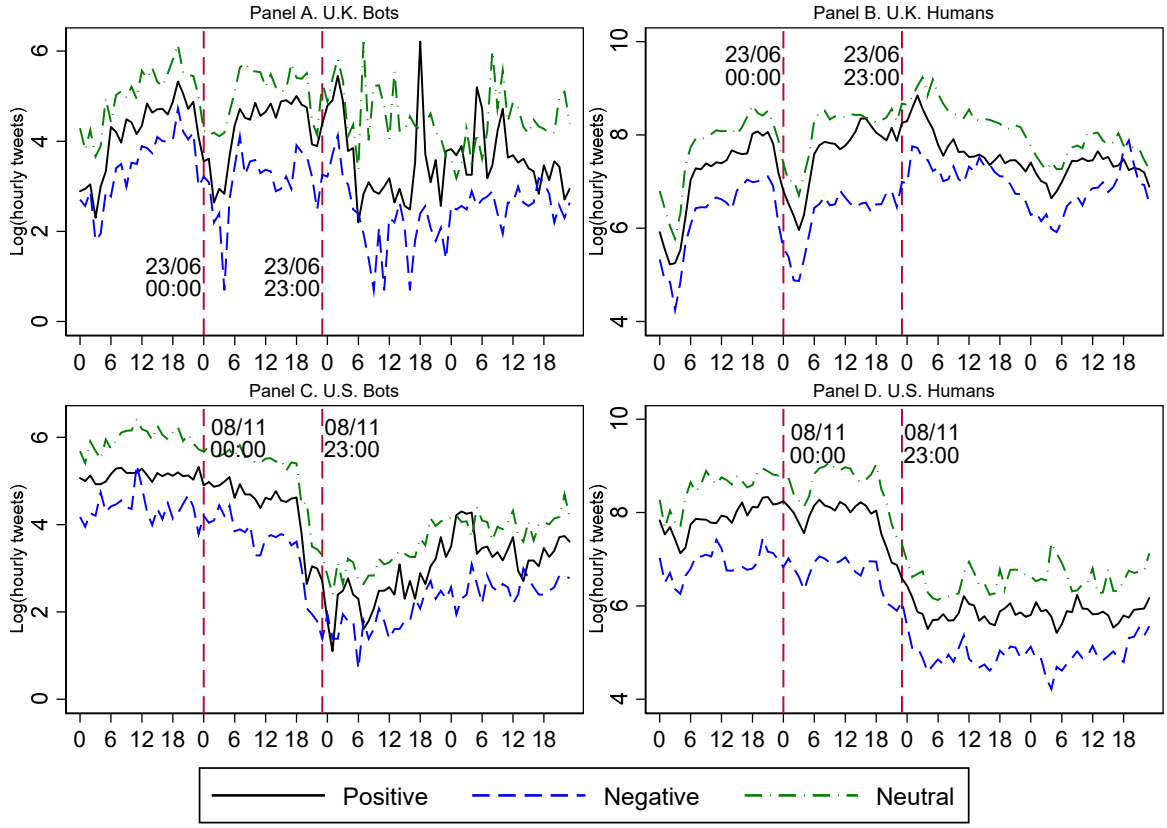
Notes: This figure shows the changes in the intensity of humans' and bots' tweeting activity by hour of the day and by day of the week.

Figure B2. Distribution of tweet sentiment



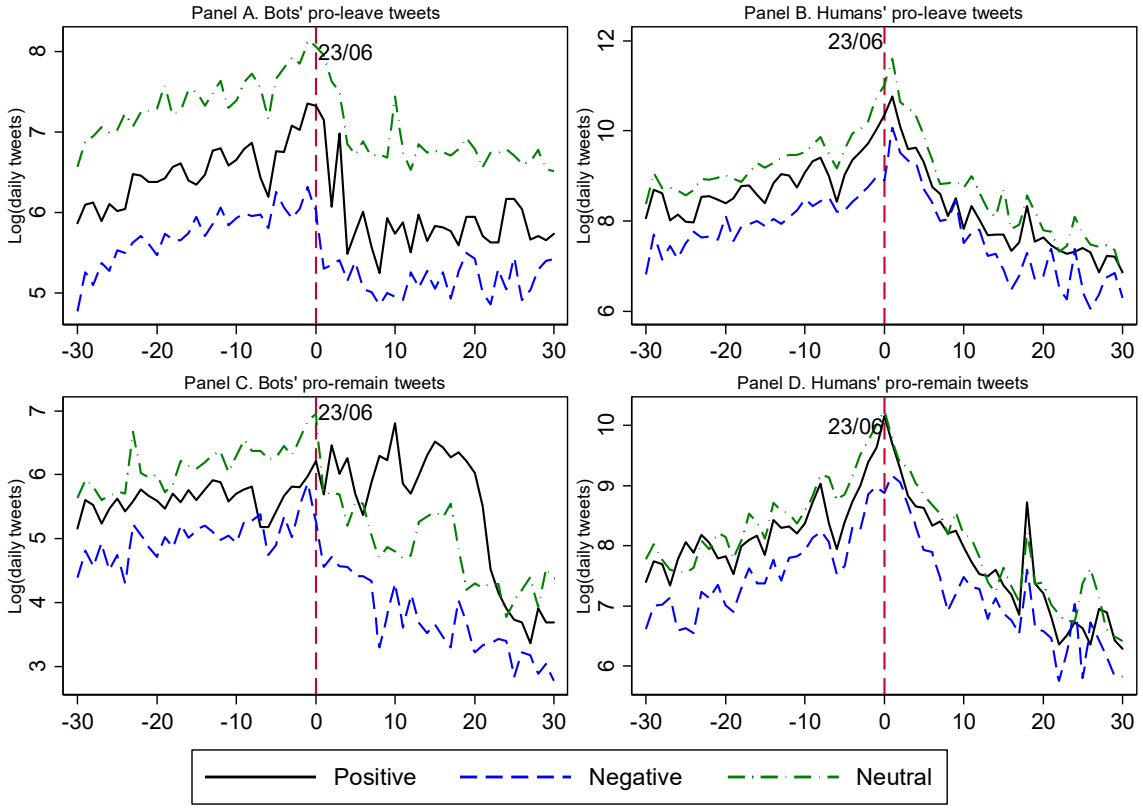
Notes: This figure shows the distribution of tweet sentiment.

Figure B3.



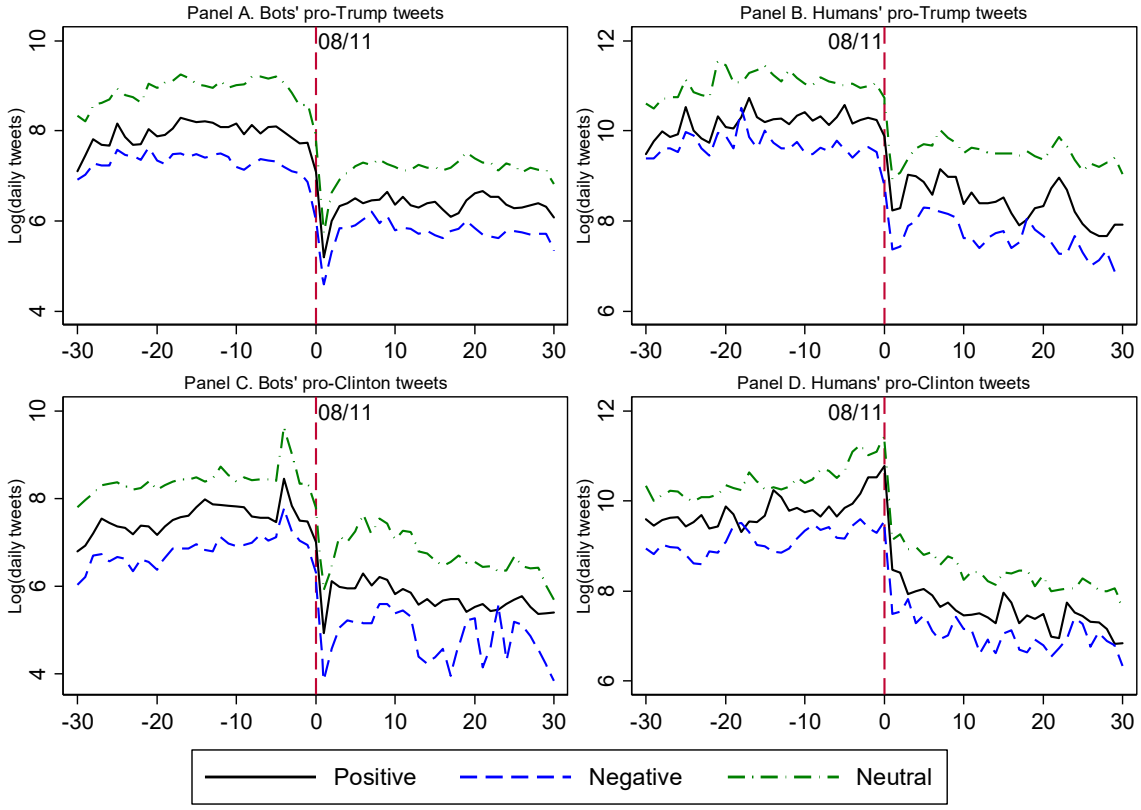
Notes: This figure shows the hourly volume of tweets around the voting days with sentiment created in humans and bots in the Brexit sample (Panels A and B) and the U.S. Election sample (Panels C and D). The dashed blue line, the solid black line, and the dashed-dotted green line show the volumes of tweets with negative, positive, and neutral sentiment, respectively.

Figure B4.



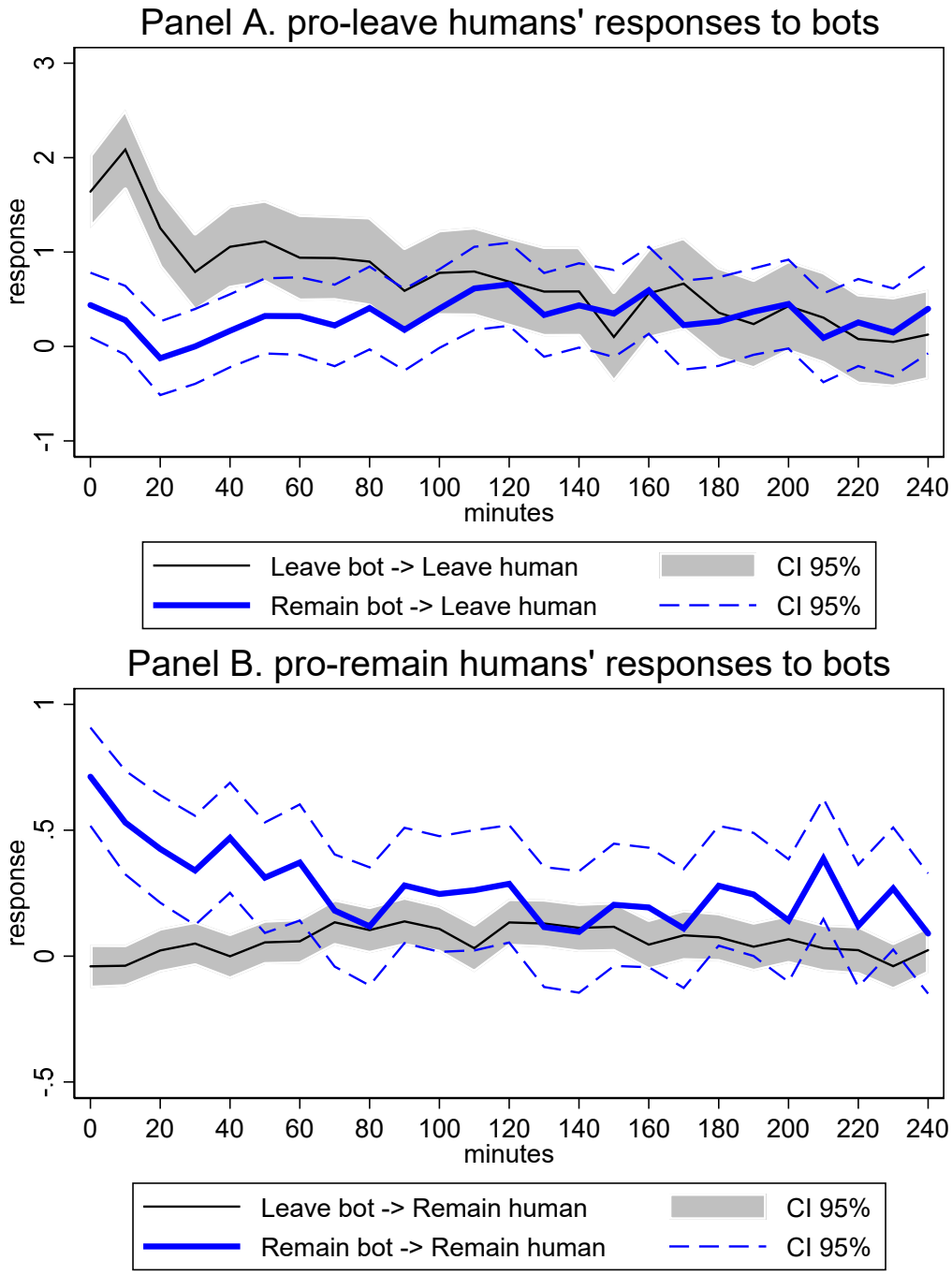
Notes: This figure shows the daily volume of tweets with sentiment created in humans and bots for pro-leave tweets (Panels A and B) and pro-remain tweets (Panels C and D). The dashed blue line, the solid black line, and the dashed-dotted green line show the volumes of tweets with negative, positive, and neutral sentiment, respectively.

Figure B5.



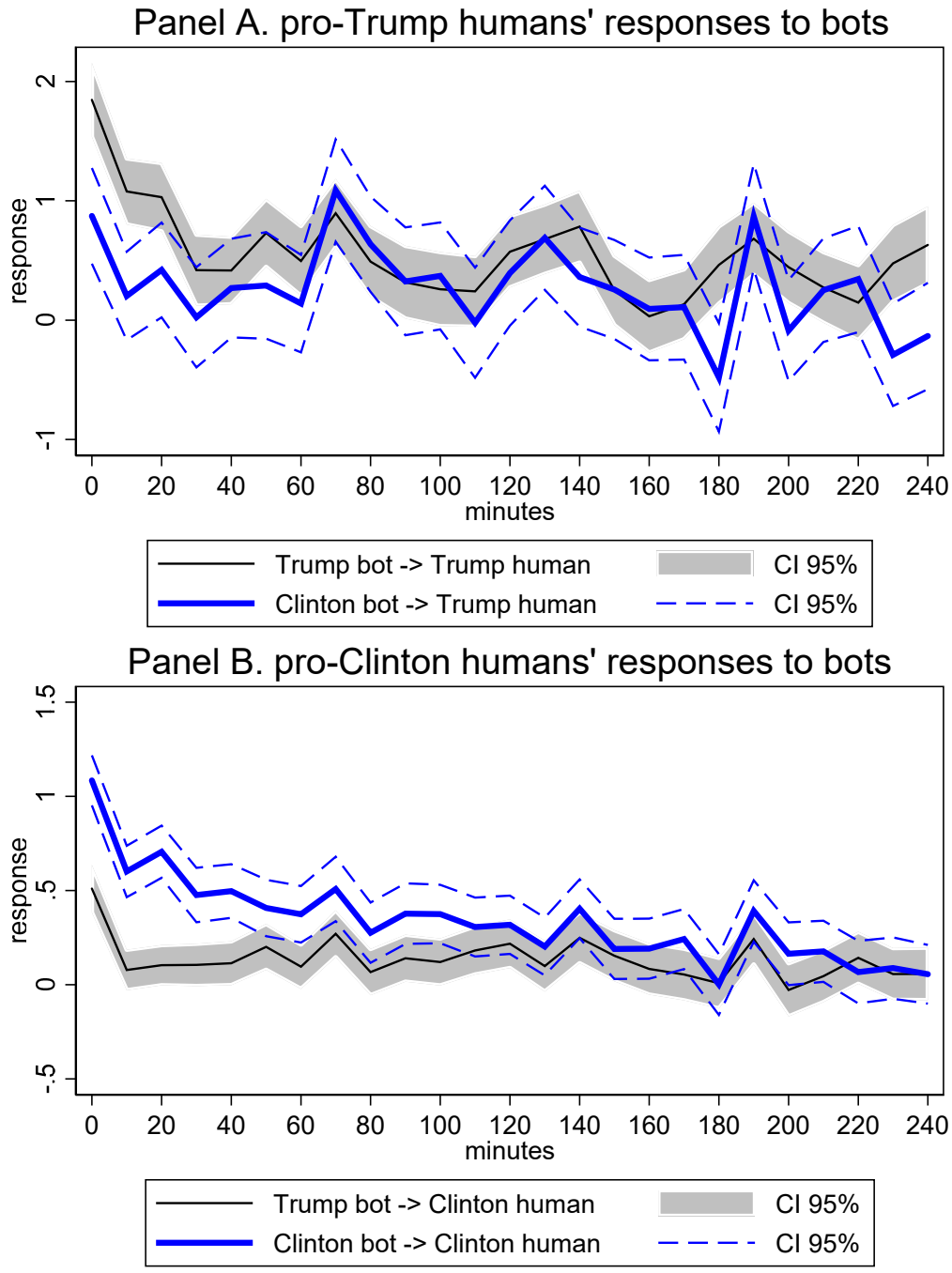
Notes: This figure shows the daily volume of tweets with sentiment created in humans and bots for pro-Trump tweets (Panels A and B) and pro-Clinton tweets (Panels C and D). The dashed blue line, the solid black line, and the dashed-dotted green line show the volumes of tweets with negative, positive, and neutral sentiment, respectively.

Figure B6. Control for news stories – U.K. Brexit



Notes: This figure reports estimate impulse responses of humans to tweets created by bots for the Brexit sample when news is controlled for. Panels A and B show the reactions of humans supporting the leave campaign and the remain campaign, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions to pro-leave bots while the thick blue line refers to the reactions to pro-remain bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-leave bots and pro-remain bots, respectively.

Figure B7. Control for news stories – U.S Election



Notes: This figure reports estimate impulse responses of humans to tweets created by bots for the 2016 U.S. Presidential Election sample when news is controlled for. Panels A and B show the reactions of humans supporting Trump and Clinton, respectively. Time (horizontal axis) is in 10-minute intervals. The thin black line refers to the reactions to pro-Trump bots while the thick blue line refers to the reactions to pro-Clinton bots. The grey shaded area and the dashed blue lines indicate 1.96 standard deviation confidence intervals for responses to pro-Trump bots and pro-Clinton bots, respectively.

Appendix C. Predictive power of public opinions on Twitter

Previous studies show that Twitter activity may have *predictive* power for electoral outcomes (e.g., Bermingham and Smeaton, 2011; O'Connor et al., 2010; Tumasjan et al., 2011; Burnap et al., 2016). To explore whether this is the case in our sample, we compare Twitter support and the actual shares of votes received by the sides of the campaigns at the regional level.

To construct the former, we use the location of Twitter users to measure how intensively a given geographical location (a state for the U.S. and a region for the U.K.) supports a given side. Takhteyev et al. (2012) documents that 75% of Twitter accounts in their large sample report geographical locations of their owners. We find that a very similar share of users reports their location in our sample. While owners may choose locations different from where they actually reside, available evidence (e.g., Takhteyev et al., 2012; Haustein and Costas, 2014) suggests that, while imperfect, this information is useful for determining geography of Twitter users.[23]

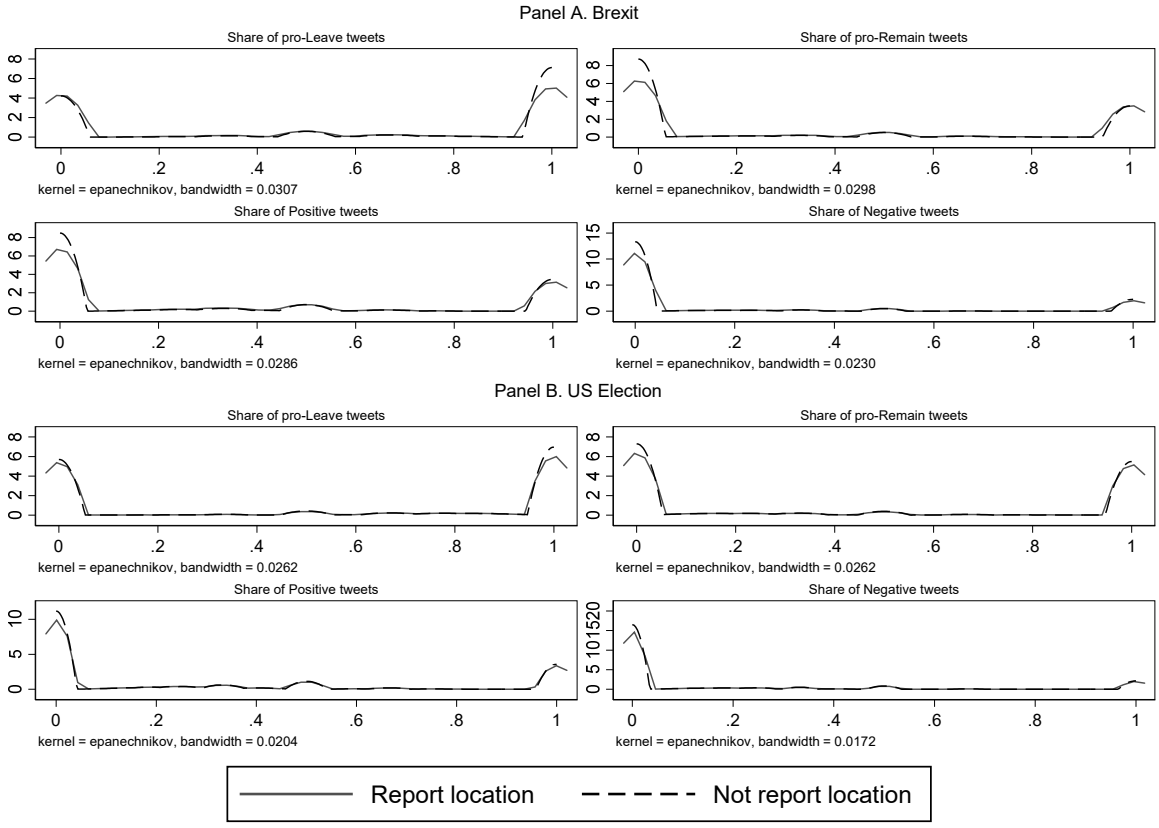
Once the location of users is established, we compute the share of pro-leave "human" tweets in total "human" tweets on the day before Referendum for the Brexit sample and the share of pro-Trump "human" tweets in total "human" tweets on the day before the vote date for the 2016 U.S. Presidential Election sample. Figure C2 shows that these shares are highly correlated with the shares of votes received by the corresponding platform.[24] To be clear, these correlations are *not causal*, and they do not measure persuasion rates in the spirit of DellaVigna and Kaplan (2007). Instead, these results only suggest that the volume of Twitter activity may be a useful *predictor* of electoral outcomes in our samples.[25]

[23] We do not observe any systematic differences in the average political slant or sentiment of tweets between users who disclose their locations and those who do not (Figure C1). Furthermore, if location were often misreported, we would be less likely to find a high correlation between locational political slant in Twitter and locational political slant in voting behaviour.

[24] Because voters could write in candidates in the U.S., the actual votes by the U.S. states are calculated using this formula: Actual vote = Votes for Trump/(Votes for Trump + Votes for Clinton).

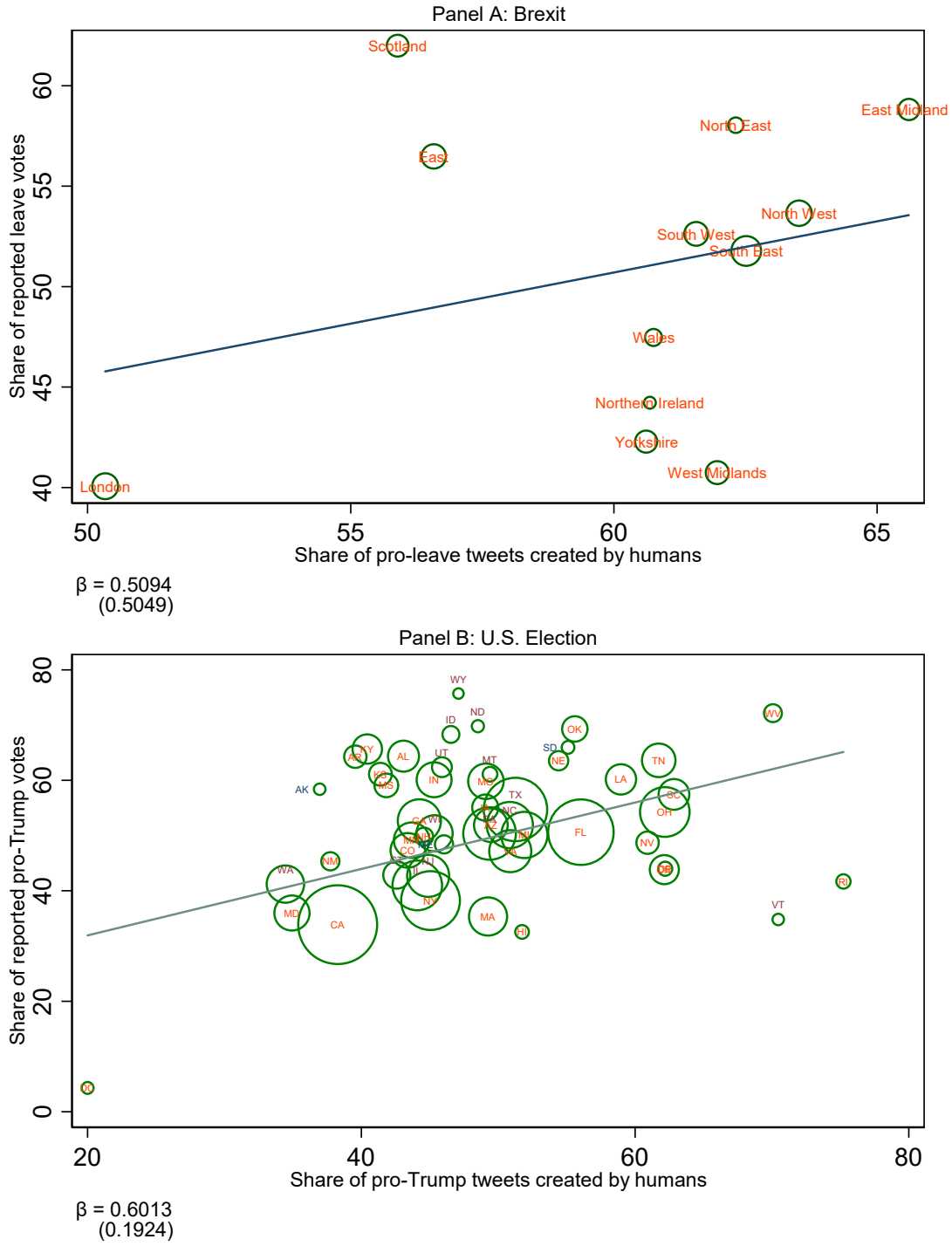
[25] We find similar results when we weight tweets with sentiment.

Figure C1. Kernel density plots based on location disclosure



Notes: This figure shows the distributions of daily share of pro-Leave, pro-Remain, Positive, and Negative tweets for Brexit sample (Panel A) as well as the distributions of daily share of pro-Trump, pro-Clinton, Positive, and Negative tweets for U.S. Election sample (Panel B). Each share is calculated at user level. The solid line shows the distributions for users who disclose their locations while the dashed line shows the distributions for users who do not disclose their locations.

Figure C2. Twitter activity and vote outcomes by geography



Notes: This figure shows the correlation between the shares of humans' pro-leave tweets and the actual vote shares by region for the Brexit sample (Panel A) and the correlation between the shares of humans' pro-Trump tweets and the actual vote shares for the 2016 U.S. Presidential Election sample (Panel B). β below each panel shows the estimated slope and standard error (in parentheses).

Appendix D. Historical contribution of bots

Our analysis suggests that bots may have tangible influence on the tweeting activity of humans. To quantify the contribution of bots' traffic to the volume of human tweets, we use the method developed by Coibion et al. (2017) to construct historical decompositions of human tweet volumes. In particular, we are interested in constructing counterfactual time series of human tweets that would have been observed in the absence of bot traffic.

To implement the Coibion et al. method, we make two departures from specification (1). First, we use innovations to bots' volume of tweets rather than the level of bots' tweets on the right-hand side of specification (1).[26] This step is necessary to have the dynamics of human tweet activity as a moving average process in terms of "bot" shocks so that we can construct a counterfactual dynamic of human tweet activity when we turn off "bot" shocks. We continue using lags of human tweets in levels (rather than in shocks). As a result, we have a combination of moving average (MA) terms (that is, current and lagged values of bot shocks) and autoregressive (AR) terms (that is, lags of human traffic supporting campaign X and lags of human traffic supporting campaign Y). To ensure that we have enough persistence in this vector ARMA representation of the stochastic process for human tweets, we increase the number of lags K from 24 to 99. We find that impulse responses based on this modification are virtually identical to the impulse responses based on specification (1). Thus, for all practical purposes, this modification does not alter our previous conclusions.

Second, we do not include dummies ψ_d for each day of a campaign. While these dummy variables are helpful to control for trends in the data, they also demean bots' traffic so that the contribution of bots to the *daily* volume of human tweet activity is zero by construction. Fortunately, we find that removing ψ_d makes little difference for the estimated impulse response and, therefore, our previous conclusions continue to apply.

Note that the dynamics of human tweets are now modelled as a system of two equations with two endogenous variables (e.g., "leave" human tweets and "remain" human tweets) driven by bot shocks (e.g., "leave" bot shocks and "remain" bot shocks) and by shocks to human tweet

[26] The shocks to bots are constructed as follows. We use specification (1) with bot traffic as the dependent variable and all lag polynomials starting with $k = 1$. We estimate this specification separately for bots supporting campaign X and for bots supporting campaign Y. The residual of each regression is interpreted as the shock to bots supporting the corresponding campaign.

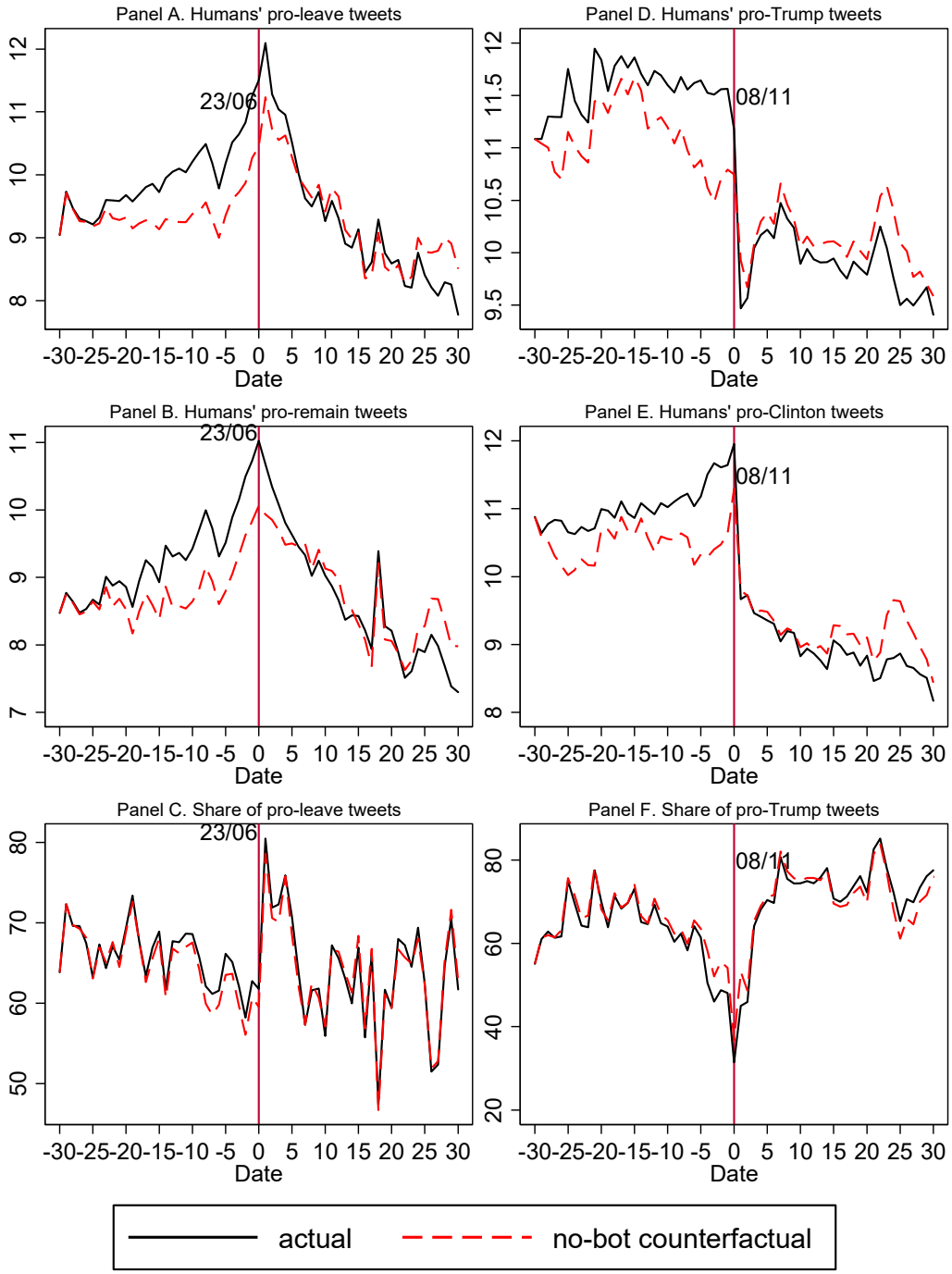
activity (e.g., $error^{(0)}_{"leave",td}$ and $error^{(0)}_{"reman",td}$). By plugging these shocks into the estimated vector ARMA process, we recover actual time series of human tweet activity. By setting bot shocks to zero, we construct counterfactual dynamics for human tweet activity when bots are not present. To make time series easier to interpret, we aggregate the generated volumes from 10-minute frequency to daily frequency.

Panel A of

Figure D1 plots times series of actual and counterfactual ("no bot") daily volume for pro-"leave" human tweets in the Brexit sample. The difference between the lines is the bots' contribution. The dynamics of the series suggest that bots had a considerable contribution to the volume of human tweeting with the largest contribution around the vote date. Panel B of the figure documents a similar pattern for pro-"remain" human tweets. Importantly, the two surges generated by bots roughly offset each other: the share of pro-"leave" human tweets in all tweets is similar for actual tweeting volume and counterfactual tweeting volume (Panel C). Specifically, the actually observed share of pro-"leave" human tweets on the day before the vote day is 62.76 percent, while the counterfactual share is 60.69 percent. This is a small absolute difference, but one should bear in mind that the Brexit outcome was decided by a small margin (the share of "leave" votes was at 51.9 percent). Our analysis in Appendix C indicates that a percentage point increase in the share of pro-"leave" tweets in total tweets is associated with a 0.54 percentage point increase in the share of actual pro-"leave" votes. While this relation is *not causal*, we can use it for *prediction*. With this interpretation in mind, we find that the difference between actual and counterfactual traffic could translate into 1.12 (=0.54×(62.76-60.69)) percentage points of actual pro-"leave" vote share. Hence, while bots nearly offset each other, the difference could have been sufficiently large to influence the outcome given how close the actual vote was.

Panels D-F replicate the analysis for the 2016 U.S. Presidential Election sample. Similar to the Brexit sample, bots appear to have a considerable contribution to the human tweet traffic. In a similar spirit, the pro-Trump and pro-Clinton human traffic generated by bots nearly offsets each other: actual and counterfactual shares of pro-Trump human tweet volume in total tweet volume are 48.09 and 54.13 percent respectively on the day before the vote date (note that pro-Clinton traffic surged days before the vote date while pro-Trump traffic was more stable so that, in our calculations, bots "helped" the Clinton campaign). But again, even this small difference could have played an important role in the outcome of these close-call elections. Specifically, our analysis in Appendix C suggests that a percentage point increase in the share of pro-Trump tweets in total tweets is associated with a 0.60 percentage point increase in the share of actual pro-Trump votes. Therefore, the observed difference between actual and counterfactual pro-Trump tweet shares suggests that 3.62 percentage points of the actual vote could be rationalized with the influence of bots.

Figure D1. Historical contribution



Notes: This figure reports the historical contribution of bots' traffic to humans' tweet volumes. Panels A, B, and C show bots' contribution to the volumes of pro-leave, pro-remain tweets and the share of pro-leave tweets, respectively. Panels D, E, and F show bots' contribution to the volumes of pro-Trump, pro-Clinton tweets, and the share of pro-Trump tweets, respectively. Time (horizontal axis) is in days. The solid black line refers to the observed human tweet activity while the dashed red line refers to human tweet activity when bots are not present.

Appendix E. More information about data collecting and screening

Appendix E1. Information on the Twitter Streaming API rate limit

According to Twitter's documentation and information provided by Twitter staff in the Twitter Developer forum, the Streaming APIs return at most 1% of the all public tweets (Firehose) at any given point of time.[27] This means that if the filter (e.g., keywords) matches less than 1% of *all* traffic on Twitter at a time, then the returned tweets are *all* public tweets that contain the keywords. However, if an event is extremely popular, the total number of public tweets containing event-related keywords might excess the 1% cap. In this case, a random sample of the filter-matched tweets (capped at 1% of Twitter traffic) is returned.

To see how this cap might affect our dataset, consider the following fictitious example. Suppose on 23rd June 2016, the volume of all public tweets created per minute was 350,000 (this is the average per-minute traffic on Twitter), making the per-minute cap 3,500 tweets per minute. At 04:00, suppose that the total number of tweets with "Brexit" keyword was 2,000. In this case, based on the filters set in Streaming API, we would get the full set of 2,000 tweets posted at 4:00 on 23/06/2016 since the volume was far lower than the 1% cap. Now suppose that at 05:01 on the same day, the volume of tweets with "Brexit" keyword reached 10,000. Since this was above the 1% cap, a random sample drawn from 10,000 tweets would be returned and the size of the sample streamed at 5:01 on 23/06/2016 would be 3,500 tweets. While Brexit and 2016 U.S. Presidential Elections were popular events, the volume of tweets related to these events was still too small to hit the cap in a way that can measurably distort our results: the daily volume of Brexit and U.S. Election tweets was in hundreds of thousands while the daily volume of all traffic on Twitter is in hundreds of millions.[28]

---

[27] For more information and discussion about Twitter Streaming API rate limit, see, for instance, https://twittercommunity.com/t/how-much-is-1/95878; https://twittercommunity.com/t/streaming-api-rate-limit/15082/4; https://twittercommunity.com/t/streaming-api-limit/93991; https://twittercommunity.com/t/current-streaming-api-sampling-rate/74864 (Accessed on 24 August 2018).

[28] Through access to Twitter Firehose, Lopez et al. (2017) are able to extract a sample of more than 30 million Brexit-related tweets covering the period from 6th January 2016 to the Referendum day, equivalent to the daily volume of approximately 175,000 tweets. This study also suggests that in the context of research that use Brexit-related tweets, Streaming API would provide similar outcomes to Firehose.

Appendix E2. Example of screening procedure

To illustrate our screening procedure, consider the following tweet posted on 26/05/2016: "RT @Nigel_Farage: I don't believe these official figures and I'm sure the real numbers are much higher. We must Leave EU and control our border https://t.co/3pr." While processing this tweet, we remove "https://t.co/3pr" and the user identifier token "@". We classify this tweet as pro-leave. Now consider another tweet posted on 10/06/2016: "This is an awesome TED talk from a couple years ago about thinking globally. Very relevant in light of #Brexit vote. https://t.co/BZeZ3fMFyA." We also remove the hashtag token and the URL link in tweet processing. Although this tweet includes references to Brexit, we do not include it in our analysis because it does not contain endorsement of any side of the debate.

Appendix Figure E1. Sample construction

	No. of tweets	
	Brexit	US Election
Original sample (raw data)	28.6 million	181.6 million
After: 1. Removing special characters in tweets' content 2. Removing tweets that contain only URL links 3. Removing non-English tweets	21.9 million	174.3 million
After 1. Classifying supporting sides 2. Removing tweets that contain keywords supporting both sides 3. Removing tweets that do not contain any keywords	 21.5 million 2.7 million	 174.1 million 7.2 million