Is political messaging persuasive or manipulative?
Twitter messages about the Supreme Court nominations

John Brehm and Laura Rossi
University of Chicago

January 21, 2020

Abstract

When we talk about politics, what we say is a mix of messages that activate deliberate thought, messages that activate automatic thought, as well as simple factual information or even simple noise. Messages that activate deliberate thought are *persuasive* messages that have the possibility of inducing long–term attitude change, and those that activate automatic thought are merely *manipulative* and tend to induce only short-term attitude change. In this paper, we draw upon three massive datasets of Twitter feeds about the nominations to the Supreme Court of Merrick Garland, Neil Gorsuch, and Brett Kavanaugh to answer the question about the balance of persuasive and manipulative messages, to identify characteristics of the "tweeters," and to identify what kind of tweeter is most likely to send what kind of message. We find that the messages during the politically contentious (if predictably futile) Garland nomination were principally manipulative messages. During the nearly inevitable Gorsuch nomination, the messages were more deliberative, and little affected by partisans, though possibly by the most extreme of bots. The Kavanaugh nomination began as inevitable, but changed in September. Those with strong feelings were very active in sending manipulative tweets when the outcome of the nominations was little in doubt, but became more deliberative when the outcome became uncertain. Manipulation occurs during political futility, but deliberation occurs when political opportunities are real.

An early version of this paper was presented at the 2017 Midwest Political Science Association annual meetings, Chicago IL, to Duke University Political Science faculty, and at the Workshop in Quantitative Methods at the University of Chicago. Thanks to Gerry Rosenberg for his consultation about the confirmation process, to Emily Thorson and James Evans for their comments on early versions of the paper. This work was completed in part with resources provided by the University of Chicago Research Computing Center.

Are political messages intended to induce longer term attitude change, or only very short term attitude change? The first mode is "persuasion," intended to actually change the mind of the recipient in some important way, perhaps by changing the evaluation of the message, or more likely by encouraging prior convictions. The second mode is "manipulation," as it is likely only to change the mind of the recipient for the relatively short term, perhaps for purposes of an imminent decision, because of the relative costs of manipulating instead of persuading, or other reasons.

The general purpose of this project is to argue for the value of studying *persuading* as opposed to *persuasion*. The persuasion literature in both political science (e.g., Mutz, Brody, and Sniderman 1996) and social psychology (e.g., Petty and Cacioppo 1986) is principally concerned with the effects of messages upon the recipient. In this broader project, we want to argue that the kinds of messages that actors themselves send are worthy of systematic study. After all, one of the core lessons from the persuasion literature is that only messages that invoke self–interest (broadly construed) would be likely to induce a longer term attitude change. But it is still possible to induce shorter term attitude change by other kinds of appeals, those that evoke more extrinsic considerations like the characteristics of the messenger or message, or even aspects of the message itself which are not especially related to the matter at hand including *ad homina,* appeals to authority, reciprocity, or consistency (among others, see Cialdini 1984).

The recent nominations of Merrick Garland, Neil Gorsuch, and Brett Kavanaugh to the Supreme Court offer a superb opportunity to consider the kinds of messages that actors send. There are very core, self–interested matters that such messages could speak to the judicial records of the three as Federal Judges speak quite a bit to the kinds of Supreme Court Justices that each might become. Among many other aspects, Garland wrote major decisions on the Oklahoma City bombing case and contributed to a decision involving second amendment rights, and Gorsuch wrote decisions on corporate law and privacy rights. But Garland was nominated by a President who is widely disliked among a portion of the public, in the last year of his presidency, while Gorsuch was nominated by another President who is also widely disliked among a very different portion of the public, in the tumultuous first year of his presidency. President Trump's nomination of Kavanaugh to replace Anthony Kennedy began as a debate about specific judicial decisons on health care and deference to the Executive branch, but shifted suddenly in early September amid charges of sexual violence. The character of the messages that might be sent about the three candidates can range from squarely deliberative to entirely peripheral.

While one might explicitly study the messages from specific political actors, in this project, we choose to analyze the messages that were sent over Twitter over the course of the three nominations. The advantage of Twitter, of course, is that *anyone* might participate. Instead of considering messages from political elites, in this project, we consider the messages from the vast population of potential Twitter users.

To this end, we undertake a three stage project. First, each of the three nominations yielded a prodigious number of tweets (65 million during Garland, 9 million during Gorsuch, and 5 million during Kavanaugh), and each requiring distinct strategies for pre-processing. Second, we employ a statistical method, Supervised Latent Dirichlet Analysis, to separate out the tweets by the nature of their content, and then from that content, produce a measures of the number of Deliberative (more likely to persuade) and Automatic (more likely to manipulate) tweets. From the second stage, we can answer the question posed in the title to the paper. Third, we analyze the characteristics of the tweeters themselves — including the probability that the tweeter is an automaton, or "bot" — and what kind of message each type of tweeter sends.

We find: that the sentiment in the Twitter series about the nominations tended to be more positive than negative; that the messages were predominantly manipulative ones during the Garland nomination, deliberative during the Gorsuch nomination, and markedly changed in form during the Kavanaugh nomination; and that partisans and bots were more likely than others to be sending manipulative messages (much of the time), but that only the most extreme of bots would have been responsible for manipulative messages in the Gorsuch and Kavanaugh nominations. Most of all, the strongest evidence here is that the tweeters sent messages that were consistent with the likelihood of eventual confirmation: when the nomination's outcome was little in doubt, tweeters with a high affect towards the nominees sent manipulative tweets; when the nomination's outcome was uncertain (for as little a time period as that would be the case), tweeters were more likely to send deliberative messages.

1 Deliberative or Automatic Messages

Political messaging can take at least two forms: "persuasive" messages or "manipulative" ones. By "persuasive" messages, we refer to those messages that have a high probability of inducing enduring attitude change. Persuasive messages are those that change the receiver's mind or further convince the receiver of a prior position (e.g., Mansbridge 1983). From the view of democratic deliberation, persuasive messages are subjectively good for the polity in that subsequent decisions by the receiver would be based on stable attitudes and a subjective evaluation of the facts. By "manipulative" messages, we refer to those messages that have a high probability of inducing ephemeral attitude change. From the view of democratic deliberation, manipulative messages are poisonous for the polity in that the receiver's short–term decisions may be unlikely to be based on stable attitudes, and instead may be influenced by largely irrelevant logical fallacies that do not hold on subsequent reflection.

These two forms of political messages are ubiquitous. Newspaper editorials, for example, might aim to change the reader's opinion about a matter of policy or a candidate by appeals to the intrinsic merits of the policy for the receiver or the reasoned differences between the candidate and the alternatives. Conversely, the editorials might influence the reader by referring to such extrinsic attributes as the policy sponsor's race or likability, the consistency of the policy with other recently passed policy, or lie outright about core facts. In an extensive analysis of newspaper editorials about the Affordable Care Act (Brehm and Tutunik 2016), we found that the overwhelming preponderance of the messages were those that could be classified as manipulative ones, by a 19:1 ratio, even by well-informed advocates.

We believe that the shift of lens from the effects of messages *on* recipients to the choices *by* senders to make certain kinds of messages is largely unique to the field, yet it can be greatly informed by the prior work. There is a rather sizable literature on persuasion effects in politics, but principally about the persuasive effects of messages to receivers in certain contexts. For example, we know something about the persuasive effects of messages received by voters on their issue preferences (e.g., Abramowitz 1978; Lanoue 1992; Yawn *et al.* 1998) which show that voters' issue preferences converge towards those of their preferred candidates *after* exposure to information. We also know something, although much more limited, about the effects of certain kinds of messages on the electorate's prospects for turnout, sense of political efficacy, and generalized trust in government (Ansolabahere and Iyengar 1995; Ansolabahere, Iyengar, and Simon 1999; Huber and Arceneaux 2007). We know that certain kinds of messages that arouse fear in the recipient move people from a state of acquiescence to greater attention to character (in the case of lower political sophistication voters) or to issues (in the case of higher political sophistication voters) (Marcus and MacKuen 1993, Brader 2005). In that vein, we know a little about the choices that candidates themselves make to "go negative" in their political advertising (namely, when the opponent has attacked earlier, when they are behind in the polls, and closer to Election Day; see Damore 2002).

Hillygus and Shields (2008) are really quite unique in considering not just the effects of certain (issue based) messages upon the potential voters, but also about the strategic choices that candidates and parties make because of the potential effects. In particular, they find that "Candidates are more likely to use divisive wedge issues when they have more information about the preferences of the voters and when they are able to narrowly target their campaign messages (p. 184)." A very strong illustration of their results would be in the evolution of campaign and party strategy with respect to the "Southern Strategy" of shifting Republican and Democratic party and candidate positions on race in the knowledge of how well the positions would "play" with

Southern voters.

Instead of a focus on the politically pertinent content of the message, this paper (and the general project) asks about the persuasive nature of the message itself.

One particular area where the persuasion literature can inform our study of persuading is by understanding the kinds of messages that are most likely to be effective. In social psychology, one of the most productive and informative frameworks has been adopted around the "Elaboration Likelihood Model" (Petty and Cacioppo 1986). In their Elaboration Likelihood Model, the message sent can be processed by the recipient in one of two paths, the "central" or "peripheral" routes to persuasion. Messages are processed under the central route to persuasion if two hurdles are crossed: does the recipient have the ability to understand the message? is the recipient motivated to receive the message? One might be quite intrigued (or highly motivated) to receive a message about a scientific breakthrough, yet lack the ability to comprehend the breakthrough. Or one might have the ability to comprehend the breakthrough, but not see the message as worth the effort to consider. If the recipient is both motivated and capable of receiving the message, then the recipient engages in a subjective consideration of the *intrinsic* merits of the message.

If the recipient is not motivated or not able to process the message, then the message may still affect the recipient's views, but does so through the peripheral route where the *extrinsic* aspects of the message matter more. Whether the receiver finds the messenger attractive or unattractive, the medium of the message to be interesting or unengaging, the context to be focusing or distracting *can* affect whether the receiver accepts or rejects the message.

Cialdini's work (1984) implies that we can further specify what might account for the peripheral process in his work on the effects of persuasive heuristics. Whether the recipient has made a prior public commitment that entails consistency on accepting or rejecting the message, whether the sender is an authority figure, whether the message asks the recipient to accept the message under conditions of the scarcity or wealth of the choices, and other heuristics may all affect the choice, and none of them involve conscious deliberation on the part of the recipient.

We present the distinction between the central and peripheral routes as if these are in stark opposition, but in practice, these are endpoints of a continuum of conscious involvement in the reception of the message. Nonetheless, there is an important consequence of the endpoints. Messages which are received under the central route involve the conscious mind into a consideration of self–interest and tend to be enduring. Messages which are received under the peripheral route are automatic, do not involve the conscious mind, and tend to be fleeting.

In the prior research on the persuasive messages about the Affordable Care Act in the editorials of the *New York Times* and *Wall Street Journal*, we found that the vast preponderance of the messages tended to be peripheral, not central, by a 19:1 ratio. The overwhelming imbalance exists despite the enormously high prominence of the policy under question, the extraordinary importance of health care in the national economy (about one–sixth to one–fifth of the national GDP), and the elite status of the writers for both papers. Whether the writer was a Nobel laureate or former Presidential staff member did not affect the balance except potentially a small *increase* in the peripheral messages over the central ones.

The imbalance is stunning. If one wanted to actually convince someone of your point of view, and you knew that certain messages had only an ephemeral effect on the recipient, why would you choose to send an ephemeral message?

At first consideration, if we saw so few deliberative (central) messages in the elite newspaper coverage of the ACA, why would we expect that picture to be any more in favor of deliberative messages with the Twitter streams about the Garland and Gorsuch nominations? After all, the newspapers only permit very able and motivated people to write in the Op-Eds, but *anyone* can post on Twitter. And a short tweet[1] would seem to beg for gratuitous pokes instead of considered thought.

[1]Twitter maintained a maximum tweet length of 140 characters during Garland and Gorsuch, but between the Gorsuch and Kavanaugh nominations increased the length of tweets up to 280 characters.

But there are plausible reasons to still regard tweets as a viable base of political communication. For one, there are major thresholds to publishing an OpEd in the *NYT* or *WSJ*, some self–imposed, some imposed by the Editorial Board. It takes a rather motivated individual to prepare a lengthy message in an OpEd, while the threshold for participation in Twitter is lower (which might mean either more deliberate *or* more automatic messages). The newspaper editorial boards have their own thresholds for publishing, though exactly what these are is elusive to us. The complexity of the Supreme Court nominations might be quite high, or quite low: it is possible to have a considered, deliberate opinions about a nominee on the basis of not particularly elaborate details such as qualifications or ideology.

Moreover, the OpEd writers in the respective newspapers speak to an audience that may not be all that representative of the general public. With the vast number of Twitter subscribers, (approximately 68 million in the US alone), open to anyone, the audience is considerably larger.

If the question is "what form does political messaging take?" the qualifiers are immediately about what? under what conditions? by whom? and to whom? And we are skeptical that there is a universal answer, nor that there is anything like a "population" of messages to sample from independent of the context. Knowing something about the balance of deliberative and automatic messages in the specific context of Twitter messages about the two specific nominees in this (one might hope) extraordinary political time is informative on its own.

1.1 What constitutes a "deliberative" or "automatic" message about the Supreme Court nominees?

We consciously set a low bar for what we would call a "deliberative" message on the basis of gleaning a few words from the Twitter topics (or explicitly set out a design strategy which favors finding persuasive content in a message). On this basis, we look to scholarship in political science about what Senators consider during the nominations process.

Cameron, Cover, and Segal (1990) model Senate confirmation of Supreme Court nominees as a function of two particularly important factors: is the nominee "well–qualified" and is the nominee "ideologically proximate" to the Senator? In the case of the Twitter messages, while it is conceivable that someone would actually identify the ABA's official assessment of the nominees' qualifications (all three are "very highly qualified"), it is far more likely that a tweet would make reference to the nominee as being "qualified." We will code all tweets that mention qualifications as being "deliberative" messages.

Ideological "proximity" is, of course, impossible to assess without knowing the ideology of the recipient, which the tweeters would not themselves know (since anyone can follow another tweeter). However, tweets which explicitly mention the ideology of the nominee (e.g., by using the words "liberal" or "conservative") would be ones that we also code as "deliberate."

Beyond these two specific categories, other literature in political science on the Senators' decisions about the nominees point to the nominees' support from interest groups. Sulfridge (1980) argues for the importance of ideology, measured with the ACA ratings of the nominee. We code any tweet that mentions explicit interest groups, or most importantly, makes use of the hashtags relevant to the interest groups as "deliberative." This means that tweets including '#2A" (pertaining to a decision by Garland involving gun rights) or '#NRA' would be coded as "deliberative."

Lastly, tweets that mention specific facts in the nominees past, especially those involving decisions by the nominees, would be coded as "deliberative." We consider tweets that mention Garland's role in the Federal court decision after the Oklahoma City bombing to be deliberative. Tweets that refer to the charges brought by Christine Blasey Ford (and other women) of sexual assault are also "deliberative" (in that the charges refer to judicial character and stability). Our coding of the hashtag '#MeToo' is much more problematic. In the sense that #MeToo indirectly refers to the charges by multiple women of sexual violence against Kavanaugh, we choose to code those tweets as "deliberative" ones. But like #NRA or #2A, the amount of reflective thought that #MeToo induces is potentially quite small.

These are *very low* bars for what we consider to be an indication of deliberative thought. Simply referring to a nominee as a "liberal" or "conservative" with no other basis does not ask for much in the way of consideration.

However, we do want to distinguish between these kinds of persuasive appeals, and those that invoke a more automatic response.

Following Cialdini (1984), attempting to persuade someone on the basis of one's authority invokes nothing other than deference to authority. In this case, we would code mentions of the prominent supporters or opponents (Obama, Trump, Clinton) as automatic, in lieu of any indications of deliberative terms.

Likewise, appeals to the consistency of behavior apply for both nominees. In the Garland nomination, the hashtag "#DoYourJob" was directly an appeal to the Senators to hold hearings (and perhaps continue to a vote). In the Gorsuch nomination, the hashtag "#Resist" was an appeal to cease consideration of Gorsuch simply on the basis of being Trump's nomineee.

In neither of these cases of appeals to authority or to consistency, the tweet does not ask for the reader to think about the nominee, but to act automatically.

We will not be able to code the modes directly from the millions of tweets, but instead will do so on the basis of common underlying (latent) topics that influence the content of the tweets. But first, how do we process the massive number of tweets prior to data analysis?

2 Assembling and cleaning the data

We began collection of the tweets for the Garland series by using the stock Twitter API (not the "firehose") and requesting 10,000 tweets that contained the word "Garland," every half hour from the moment that President Obama announced Merrick Garland on 16 March 2016 as his nominee to the Supreme Court, discarding all retweets, obtaining roughly 3,000 tweets per half hour request, and concluding when Garland's nomination was no longer active (on 3 January 2017 when the 115th session of the U.S. Congress opened). This effort resulted in a prodigious datafile of 65,268,655 tweets.

We began a similar collection of tweets for the Gorsuch series by a similar procedure: requesting 10,000 tweets every half hour on 1 February 2017 (once Trump announced the nominee) that contained the word "Gorsuch," discarding all retweets, and obtaining about 3,000 tweets. By the time of Gorsuch's confirmation by the Senate on 7 April 2017, we gathered 9,148,845 tweets. The collection of the Kavanaugh series followed in a similar fashion, beginning on 9 July 2018 during Trump's news conference, and concluding with the final Senate vote to confirm on 6 October 2018, result in 3,411,129 tweets.

The first significant task required that we clean the massive number of tweets down to only those tweets that pertained to the respective nominees. The most significant challenge is that there are, of course, many tweets that use the word "Garland" that have nothing at all to do with Merrick Garland.[2] We also excluded tweets that were not from the United States (by study design), or were exclusively in a language other than English, or used a non–English character set. We translated the non-English characters to English equivalents when the majority of the tweet itself was in English.[3]

After purging the non–Merrick Garland observations, we reduced the sample by more than half, down to 25 million tweets.

[2]A more restrictive search, say by limiting to the full name, or to various hashtags (e.g., #SCOTUS, #Garland) would not have been successful. Not everyone knows to use certain tags (e.g., #SCOTUS), and other tags would not have been sufficiently restrictive (many non-Merrick Garland tweets used the hashtag #Garland). Among the false tags: People who aren't Merrick Garland (e.g., Judy Garland, Garland Jeffreys, Connor Garland, Travis Garland); People who publicly imitated Judy Garland (as did the singer Arianna Grande), or referenced Supreme Court justices in a comedy sketch on SNL; Places that include Garland (e.g., Garland, TX or Garland County, AR); Facilities located in one of these places (e.g., a Sonic restaurant or animal shelter in Garland TX, or a police blotter reported from Garland County, AR); Facilities located anywhere, including Chile, that included "Garland" in the name; Fictional shows that had characters named "Garland"; Holiday decorations.

[3]This was of special consideration during the recent presidential campaign when the rights of Spanish-speaking people in the US was an important issue, and who might be expected to express a view on Twitter.

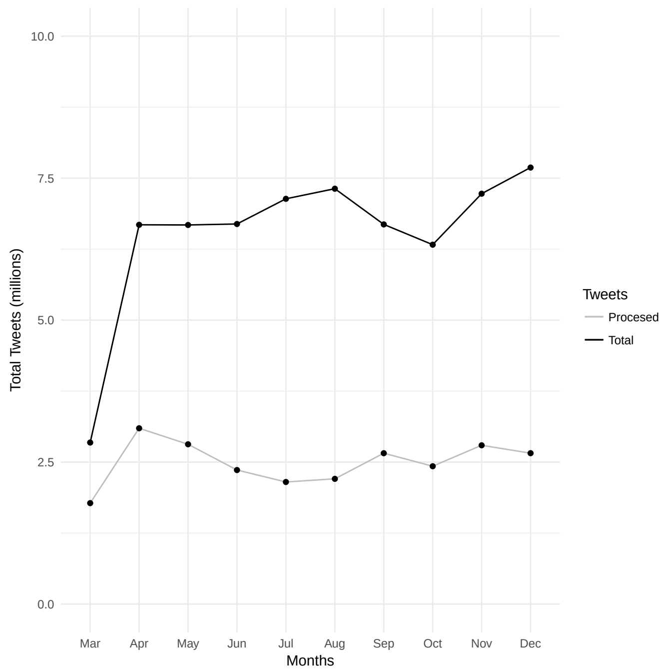

Figure 1: Tweets mentioning “Garland,” Mar–Dec 2016

Figure 1 displays the total volume of tweets using the word “Garland,” as well as those that were remaining after processing (and serve as input for the analysis). While the total number of tweets in March is lower due to the shortened observation period (half the month), what is immediately apparent is that not only was the total level of tweets accumulated consistently high, with few fluctuations, but that the number of tweets remaining after the initial cleaning is also quite high and stable.

The issues for the initial cleaning were considerably less challenging when removing those tweets that used the word “Gorsuch.” Neil Gorsuch’s mother, Anne Gorsuch Burford, was a nominee for the EPA Director in the Reagan administration, and who eventually stepped aside due to political controversy. We did not exclude any mentions of her name in the tweets since her experience as an EPA nominee was formative for Neil Gorsuch’s own experience (and may well be formative for the nature of many of the tweets). Nonetheless, we restricted the analysis to those tweets that were sent from the US, were primarily in English, and translated to English equivalents if the bulk of the tweet was written in English.

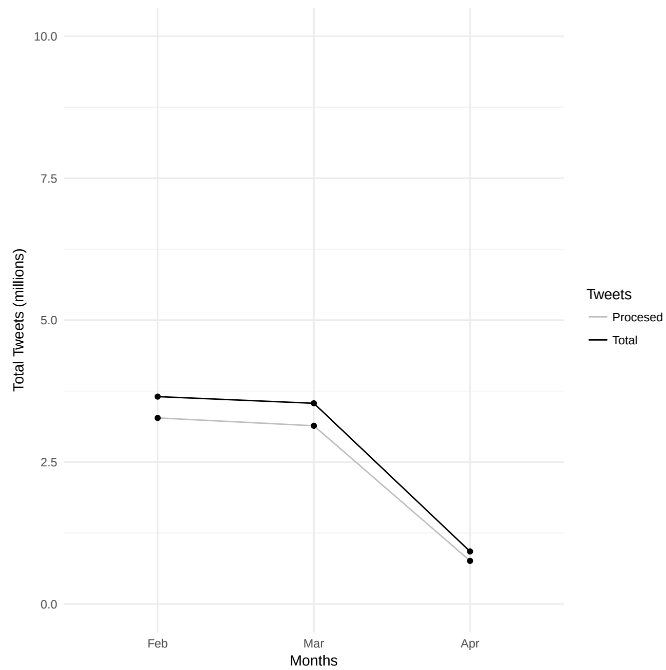

Figure 2: Tweets mentioning "Gorsuch," Feb–Mar 2017

We excluded a trending topics item "Trendinalia" that was pertinent for Gorsuch (but oddly not Garland due to a policy change at Twitter). We restricted by GPS coordinates to the US. We also excluded a few place names that included Gorsuch (e.g., Gorsuch Ave).

While the total number of tweets was clearly considerably larger for the Garland nomination, that nomination both ran for five times longer *and* included an exceptional number of cases that mentioned a non–Merrick "Garland." The monthly average after reducing the cases down to the working set was about 2.5 million tweets for Garland and 3.2 million tweets for Gorsuch.

Despite the controversy during the Kavanaugh nomination, the number of tweets during the three-plus months of the Kavanaugh series is sharply lower Figure 3. Instead of just over 2.5 million tweets per month (as during Garland and Gorsuch), the number of tweets per month during Kavanaugh drops to under a million for each of July and August, and rises to 1.3 million only in September. The cleaning process for Kavanaugh did not reduce the number by a noticeable amount (only a few thousand in each month).

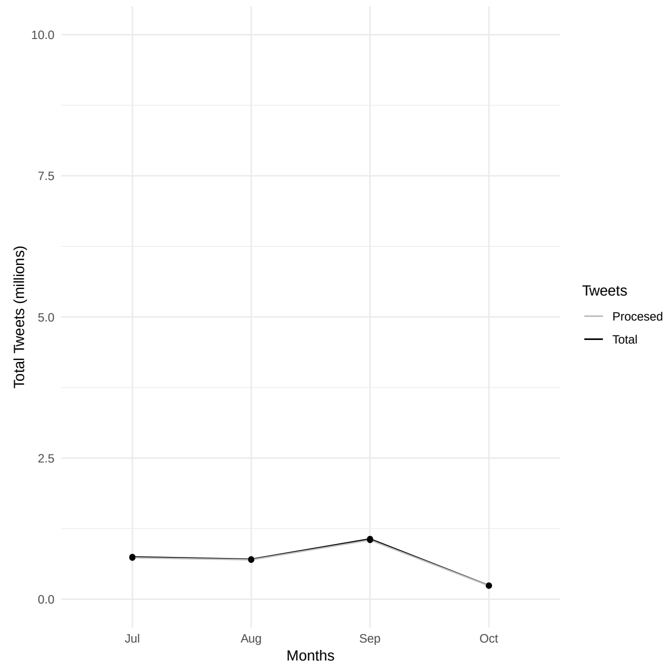

Figure 3: Tweets mentioning “Kavanaugh,” Jul–Oct 2018

One possibility for the decline in the volume of tweets could be that Twitter became no longer as popular a mode of political messaging. But what is known is that Twitter itself changed its dissemination process in two significant ways. In 2017, Twitter came under charges of lax attention to tweeting by “bots,” especially for purposes of influencing the 2016 elections. Twitter moved to purge tens of millions of twitter accounts after identifying the accounts because the accounts themselves were human. Secondly, in the year between the Gorsuch confirmation and the start of the Kavanaugh nomination, Twitter doubled the length of the tweets from 140 characters to 280. One effect, of course, is that each tweet itself contains more text, but another effect is that for those who are sending chains of connected tweets, the number of tweets could have cut by as much as half.

3 Analysis

The direct listing of the most popular words in a Twitter stream about the nominees, even in a particular month, is not especially informative, and prohibitively bulky. We need a more structured approach.

The first stage of our analysis of the tweets combines three separate techniques. In general, we will be using a "Supervised Latent Dirichlet Analysis" (sLDA) to employ a response variable (explained below), texts, and words, to generate predictions of topics. Here, the "documents" refers to the specific tweets, the "words" to the cleaned words within the tweets, and "topics" to be the general themes of discussion, which we will do on a monthly basis.[4] We then assign by hand labels to each of the topics. Last, we use the probability assigned to each topic to map to the specific predictions of whether a tweet is deliberative, automatic, other (simply relaying factual information), or not applicable. The three stages then let us simply tally the balance of deliberative vs. automatic tweets.

We obtain an initial estimate of the "response" variable (y) by analyzing the sentiment of the text using a word-matching procedure, here implemented in the `R` package, `tidytext`. We compute the sentiment of the tweet on the basis of the number of positive, neutral, and negative words, using Microsoft's "bing" dictionary.

We do note some significant issues with the word-matching approach. Sentences contain double negatives, which count as two negative words in the sequence, not as cancelling one another out.[5]. The word-matching approach does not consider larger context.[6]

[4] It is, of course, feasible to conduct the analysis on weekly or even daily bases, though these presented computational challenges in practice.

[5] The same holds for double positives, but other than the apocryphal "Yeah, yeah," we are harder pressed to come up with combinations of multiple positive words which cancel out the meaning.

[6] There are alternatives to the algorithm in `tidytext`: the particular word-matching algorithm in `RSentiment` is painfully slow: a typical dataset for a single month of 3 million tweets may take well over 12 hours to estimate sentiment on fast 64 processor machines.

The structure of the problem is that a set of documents, D, share a set of topics, but that any one D uses a mix of the topics, θ. Each document D_n is composed of words, w_n. But each w_n is related by a probability β to each topic. Our challenge is to estimate the associations between the words, documents (tweets), and topics, given some fixed number of documents, N, and initial estimates (from the sentiments), y.

Given some fixed number of topics, the algorithm for the sLDA is deceptively simple, but computationally quite demanding. Following Blei and McAuliffe (2010), the stages are:

1. Draw topic proportions $\theta|\alpha \sim \text{Dirichlet}(\alpha)$
2. For each word
   (a) Draw topic assignment $z_n|\theta \sim \text{Multinomial}(\theta)$
   (b) Draw word $w_n|z_n, \beta_{1:K} \sim \text{Multinomial}(\beta_{z_n})$
3. Draw response variable $y|z_{1:N}, \eta, \delta \sim \text{GLM}(\bar{z}, \eta, \delta)$.

(The GLM framework would permit many usages, including the probit structure here).

The sLDA can be computed in a modest length of time (2–3 hours per month under analysis, or 20–30 total hours for the 10 months of Garland data, 4–6 total hours for the 2 months of Gorsuch data, and 2 total hours for the Kavanaugh data using an expectation maximization algorithm, which is implemented in the `R` package `lda` as `slda.em`. The routine will produce a list of assignments of the topics to the tweets, which vary by each tweet (i.e., some tweets indicate a larger set of topics than others). We estimate the model for 20 topics, producing a list of the 10 most probable words per topic, for each of the 10 months of the Garland nomination, the 2.5 months of the Gorsuch nomination, and 3.5 months for the Kavanaugh nomination.

Lastly, we calculate the matching of the assignments to topics, and then to our ultimate multinomial dependent variable of interest, the mode of persuasion (m_n), which takes the form

(*D*eliberative, *A*utomatic, *O*ther, *N*A). We conduct the final coding stage by separately assigning each of the 20 topics identified per month to the four categories, and adopting the following assignment rules:

1. If both coders agree that a topic is *D,A,O,N*, we use the code;
2. The code of *D* takes precedence over *A,O,N.*;
3. The codes of *D* or *A* take precedence over *O,N*;
4. The codes of *D,A,O* take precedence over *N*;
5. If one coder did not assign a code (due to uncertainty about the topic) and the other did, we use the other's code;
6. Any remaining disagreement is discussed between the two coders.

The result is that we have, ultimately, a set of matches between topics and codes.

From this topic to persuasion mode code, we compute our assignment of the tweets to topics. For each tweet, the sLDA procedure provides estimates of probabilities that a tweet may be assigned to any particular topic. The product of a tweet's assignment to a topic and the assignment of the topic to codes yields the probability that each tweet is *D,A,O,N.* Lastly, we compute the predicted assignment on the basis of the largest probability that a tweet is of a particular topic, following a similar assignment as the coding rules above (e.g., when there is a majority in one category, *D,A,O,N,* that category is the assignment, when there is a dispute, we resolve in favor of *D.*) The procedure here is designed to favor finding that a tweet indicates deliberation over any other mode, deliberation or automatic behavior over the others.

4 Who sends what kind of message?

We want to consider four separate questions. What is the sentiment of the tweets about the nominees over the duration of the months that the nominations was active (i.e., were the tweets more likely to be negative, positive or both? What kinds of topics were the tweeters writing about in both cases (or in the latent variable idea, what kinds of topics were the tweeters expressing)? What is the predicted allotment of the tweets to the four relevant categories (deliberative, automatic, other, and not applicable)? Lastly, who is most likely to send what kind of tweet?

4.1 Sentiment by month

While the sentiment scores serve as the initial assignments, z_n, the sLDA procedure can provide us with a more continuous assessment of the actual sentiments of the tweets. What the sLDA would then allow us to consider is whether the tweet as a whole, judged in the context of all the tweets, has more positive or negative wording in the tweets. We can do this separately for Merrick Garland, Neil Gorsuch, and Brett Kavanaugh.

In both cases, we should remember that the word–level sentiment analysis can only assess whether the *words* used are negative, neutral, or positive ones. Word–level sentiment analysis is not sophisticated enough to comprehend much about the sentiment of the tweet *as a whole.* It remains an open question as to whether a more sophisticated analysis of the sentiment in a tweet as a whole is all that different from the words, and likewise an open question as to what would potentially affect a reader, the negative or positive *words* or the negative or positive *tweets.*

We consider the Garland nomination first. Figure 4 displays the kernel densities of the predicted sentiment (the "assignment" in the sLDA model) for each month of the Garland nomination. The negative tweets (by the prior sentiment score) are in the very light gray, the positive tweets are in the dark gray, and the neutral tweets are in the middle gray. We observe two points.

First, the tweets about Garland were polarized, but not consistently positive or negative. In every one of the ten months of tweets about Garland, a sizable fraction of the tweets employed more positive language than negative or neutral words: every month has a distinct mode of positive tweets (in the light gray). But a sizable fraction in most of the months were negative, though

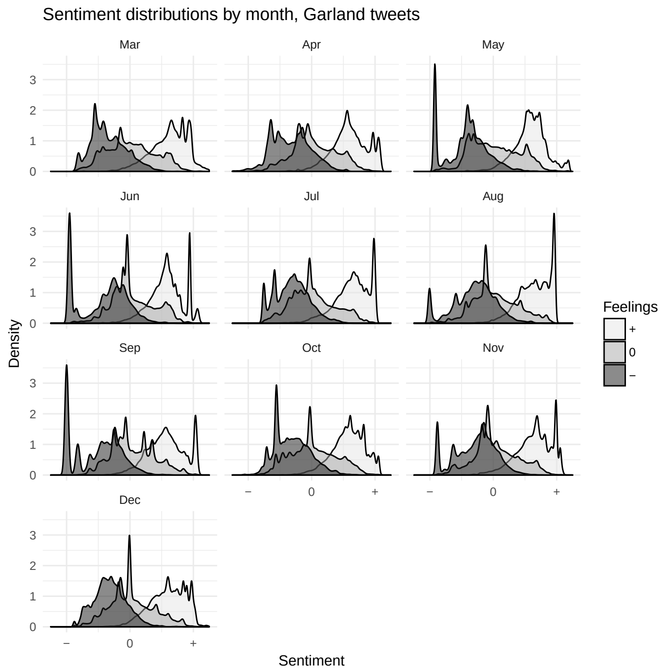

Figure 4: Changing sentiment scores in the tweets about (Merrick) Garland

not as enthusiastically negative as the tweets were positive.[7]

Second, we note that there were moments where both strongly negative and strongly positive tweets were quite common. In May, June, and September, there is a pronounced spike of negative tweets at the extreme of the scale. In June, July, and August, there is a corresponding spike of positive tweets. Some fraction of the twitter messages were uniformly negative or positive in what was conveyed, but by no means representative of the majority of the tweets themselves.

Turning to the second nomination Figure 5 presents the kernel by estimates for the predicted sentiment in the tweets about Neil Gorsuch. The same two points observed about the Garland tweets apply here: the tweets are for the most part polarized, but the mass over the positive tweets is slightly stronger than the mass of negative tweets; secondly, there is a sizable minority of tweets which were pronouncedly extreme (in this case negative, and just for February).

[7] We do want to be careful about pressing the latter point, as the choice of the "dictionary" of the sentiment scores can matter.

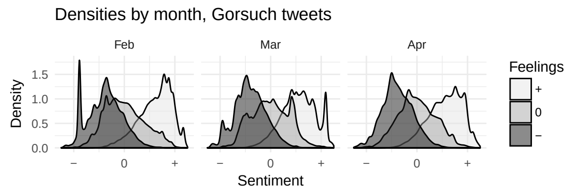

Figure 5: Changing sentiment scores in the tweets about (Neil) Gorsuch

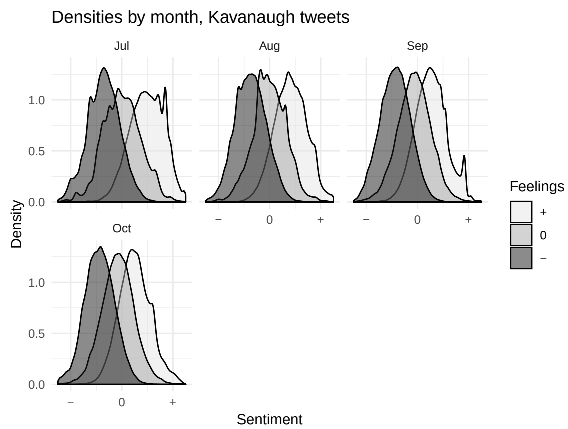

Figure 6: Changing sentiment scores in the tweets about (Brett) Kavanaugh

The sentiment scores in the Kavanaugh tweets were much less polarized than the Garland and Gorsuch series (Figure 6). Although the sentiments were slightly more polarized in July (the positive mode is high, but below the maximum; the negative mode is closer to neutral), it would be a stretch to regard these tweets as charged in the sentiment in the same way as either Garland or Gorsuch.

Across the now seventeen months worth of tweets in the three separate nominations, the general message might be this: language in the tweets can be as polarized as the polity itself, but that it is much more likely that tweeters spoke in positive terms about the nominees, even though those nominees were quite different politically. Even in highly charged months of the Kavanaugh hearings, the sentiments captured in the tweets were more neutral than polar.

4.2 What were the tweeters saying in each month?

More important than the general mood of the tweet, sLDA will permit us to discover the general topics that the tweets convey. (Or more properly, the topics themselves determine the words that comprise each tweet). While we do not wish to present all twelve displays of the distribution of sentiments by each of the twenty topics among the three different nominees, we do think it is instructive to examine how the discussion about the three nominees began.[8] (The full list of topics in all three sLDA models for each month of the analysis is available in an off-line appendix.)

We present the topic list as generated by the sLDA analysis of the first month of Tweeting about Merrick Garland as a bar chart (Figure 7). The y-axis is truncated to show only the first thirty characters of what were a full ten word list, but even so, the truncated list does give a general impression about the general discussion.

In keeping with our report of the kernel density plots about sentiment (Fig. 4), the tweeters were balanced in their language by valence, if not by the particular value. Some of the topics were more clearly informative (two of these mention Senator Kirk's decision to meet with Merrick Garland, the first Republican Senator to break with the party's position that no one would meet with the nominee).[9] Some of the topics reflected language that we considered deliberative, referring to ideology (liberal) or a policy position (guns, nature of criminal rulings). Others were tweets that only mentioned prominent politicians who were not Garland himself (Obama, Biden, McConnell, Kasich), which we will eventually code as automatic appeals (referencing an association, not the ideology, qualifications, or interest group support).

Figure 8 shows the sLDA Topics for the first month of Twitter discussion about Neil Gorsuch

[8] The topics vary in each month of the analysis, and quite possibly vary more frequently than that. Our hand–coded analysis of the equivalently long debate over the Affordable Care Act (Brehm and Tutunik 2016) showed a dramatically shifting discourse that strayed from positive cases involving coverage, extension of benefits, consistency with the general expansion of the Welfare state, through negative versions of the very same points, in language that invoked "death panels," benefits for undocumented residents, coverage on abortion, and more.

[9] Kirk was running in a very tight re–election to the Senate from the liberal state of Illinois, and ultimately lost badly. Meeting with Garland may have been a stab at trying to indicate a break from the party in general.

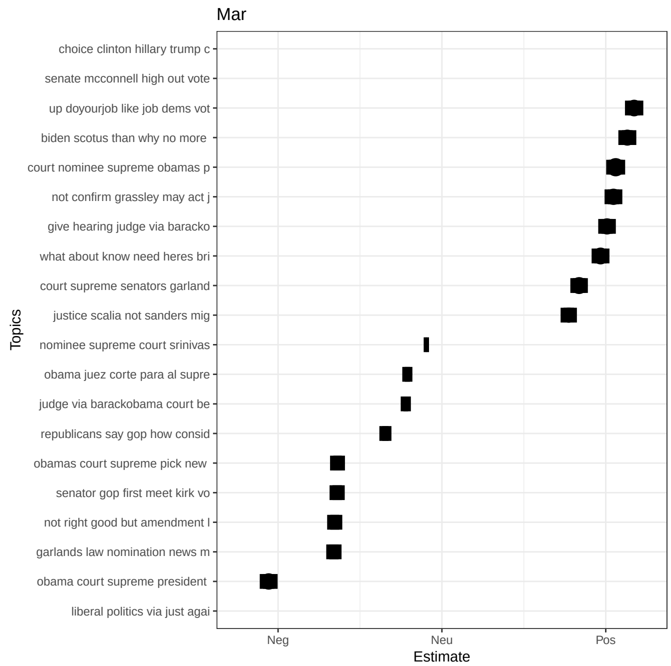

Figure 7: Bar plots of the topics and sentiments in tweets about (Merrick) Garland, Mar 2016

(Feb 2017). Again, the majority of the topics use negative words (per the predicted sentiment measure), and *nothing* is really all that positive (even those topics on the positive side of neutral are closer to neutral than positive). Seven of the topics include "trump" as the first word, which we would consider as automatic appeals, in lieu of content information about Gorsuch's ideology or interest group support. Some of these are likewise probably deliberative ones that explicitly mention ideology ("conservative"). Some appear to be more informational than either deliberative or automatic (words that reflect the news to be conveyed in the tweet).

Figure 9 displays the sLDA Topics for the first month of the Kavanaugh twitter stream (July 2018). As with the topics caught in the first months of both the Garland and Gorsuch series, the topics in the Kavanaugh series are quite polar. Recall that the debate in the Kavanaugh series would have involved mentions of Roe v Wade, potentially some mention of unusually high debt that surfaced in the FBI investigation, and the prerogatives of the Trump and the GOP, all of which is apparent. (The lines in the plot without visible dots have scores that are beyond the plotting range, or that certain topics were highly positive or negative). Even if the sentiment

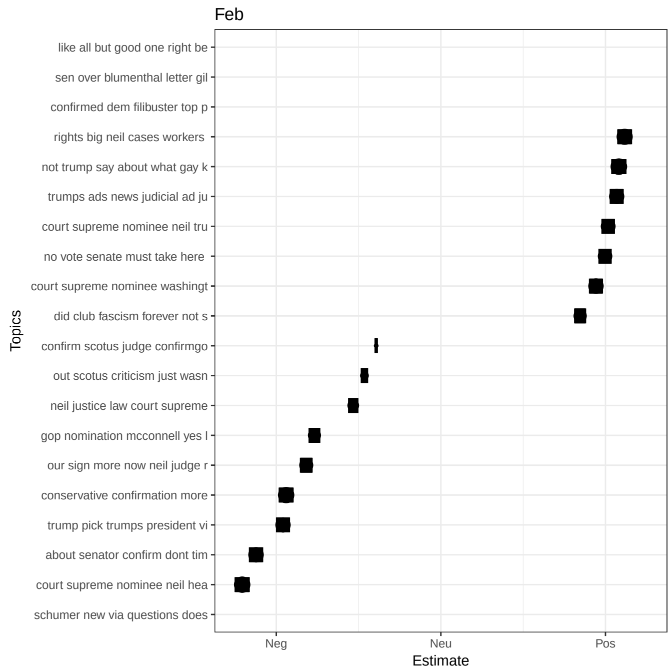

Figure 8: Bar plots of the topics and sentiments in tweets about (Neil) Gorsuch, Feb 2017

scores as a whole did not show a great deal of polarization, the language of certain topics is very much outside expected bounds.

While these are illustrative of the kinds of tweeting that happened in the first month of discussion about the nominees, the ultimate question is not about the topics, but about how we translate these topics into our four categories: deliberative, automatic, other, or not applicable.[10] Recall that our coding rules that are both restrictive about what we consider to be a deliberative tweet (limited to ideology, qualifications, or interest group support), reference some of Cialdini's (1984) categories in coding automatic tweets (especially liking, appeals to authority, and consistency), but more importantly emphasize that we would favor coding deliberative tweets over automatic ones.

[10]The last category is especially an issue for us with the Garland nomination, especially in certain months of our analysis

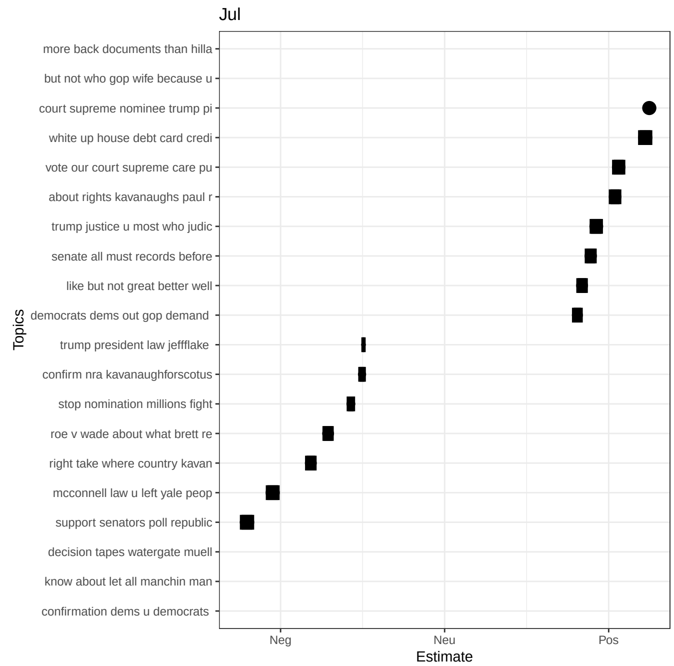

Figure 9: Bar plots of the topics and sentiments in tweets about (Brett) Kavanaugh, Jul 2018

4.3 Allotment over the four categories

The question posed in the title to this paper was whether political messaging about the nominees was persuasive (deliberated messaging) or manipulative (automatic messaging). While there are other, big questions that we want to answer with the Twitter data, for now the answer is largely unambiguous: the messaging about the nominees is contingent on the inevitability of the nominee, though there were exceptional months, and that the imbalance was nowhere near the imbalance we found on a coding of the newspaper OpEds.

We emphasize that the coding rules for the translation of the topics into modes of persuasion were explicit and in most ways favored finding deliberative messages. We coded the topic of the tweets as deliberative if it invoked one of the three criteria in the Senators' considerations (ideology, qualifications, and interest group support) or explicitly invoke self–interest. We *only* coded a topic as automatic if the message invoking authority (naming a major political figure, chiefly the Presidents) *without* also mentioning ideology, qualifications or interest group support.

We coded a message as automatic if it invoked an explicit hashtag that was part of a broader appeal to act in a consistent way. Further, when there was a choice between coding a topic as deliberative or otherwise, we chose to code the topic as deliberative. (We likewise continued the preference for automatic appeals over other or "not applicable" appeals, and for other appeals over "not applicable").

Figure 10 displays the count (in millions) of the number of tweets about Merrick Garland per mode. The two particularly relevant lines are the solid line (automatic messages) and the dashed line (deliberative messages). With the exception of the very last month, more of the tweets were automatic ones, and substantially so. In the very last month (with far fewer tweets), the number of automatic tweets falls sharply, and below the number of deliberative ones. There are also a lot of messages that fell into the "Other" category, and these were typically news reports about the nomination, not messages that appear to aim to change the view of the reader. At the same time, there were also a lot of messages that were not applicable to the confirmation at all, which is a reflection of the amount of dross remaining even after the significant cleaning of non-Merrick Garland references to "Garland."

Figure 11 replicates the analysis for the tweets about Neil Gorsuch. Here, the pattern is nearly the opposite from that for the Garland series: Most of the messages in the first two months were deliberative ones, only falling below the number of messages that were automatic in the very last month. Still, a sizable number of the tweets were automatic ones, between half a million and a million such tweets. By contrast, the number of "other" tweets is well below the number of both automatic and deliberative tweets, and the number of "not applicable" tweets is under 200,000.

There are significant differences in the character of the kind of tweet in the Twitter stream between the first two nominees. When the Garland nomination was active, the vast preponderance of the tweets were automatic ones: in all but the last month, not only was the number of automatic tweets more than the number of deliberative tweets, but sometimes by more than two to one (or an additional one million or so tweets, depending on the month). It is only in the

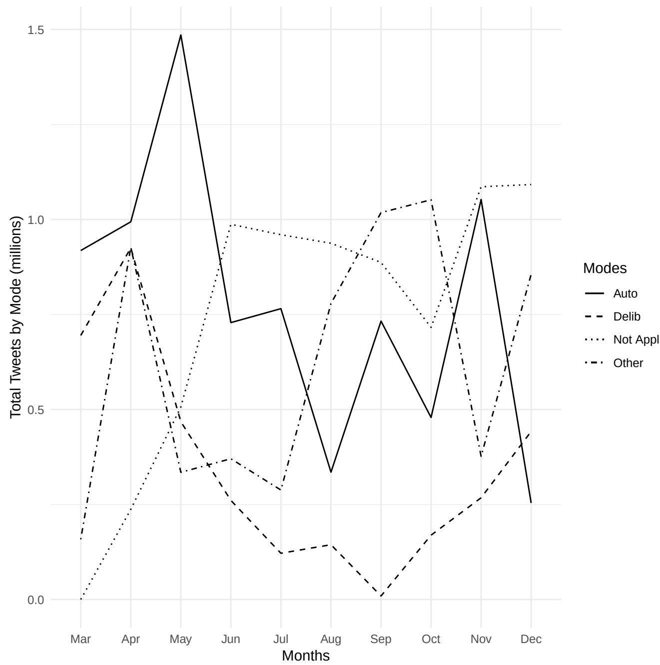

Figure 10: How many tweets about Garland per mode? (millions)

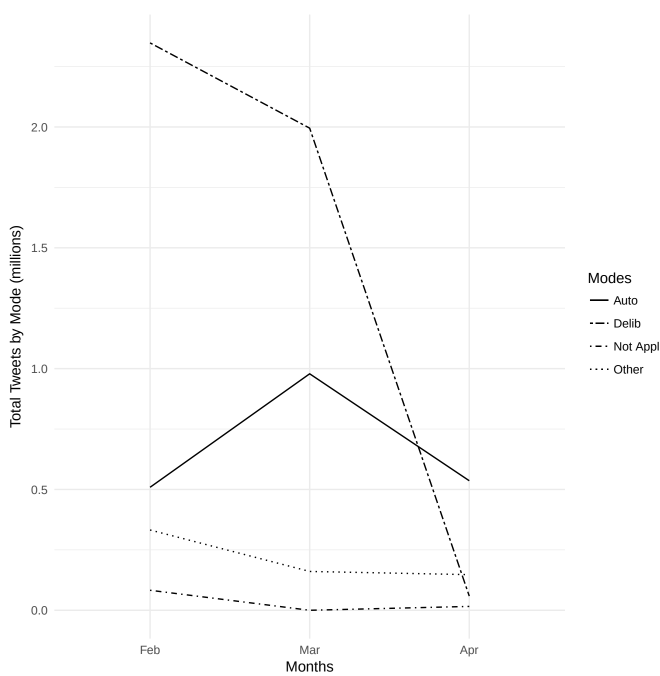

Figure 11: How many tweets about Gorsuch per mode? (millions)

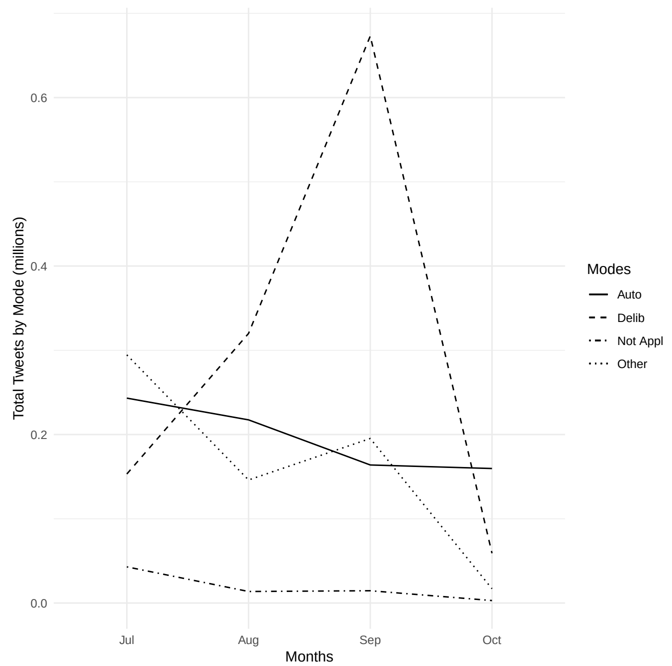

Figure 12: How many tweets about Kavanaugh per mode? (millions)

very last month of the Garland nomination that the number of deliberative tweets exceeds the number of automatic tweets. When the Gorsuch nomination was active, the picture reverses, and the number of deliberative tweets swamps the number of automatic tweets (which are still numerous), again until the very last month when the balance of deliberative tweets and automatic tweets reverses (this time, with more automatic tweets).

Why might the general character of tweets in the Garland nominations be "automatic" (or manipulative), while the general character of the Gorsuch tweets be "deliberative" (or persuasive)? Why would the character of the tweets shifted from manipulative to deliberative? There is a wealth of plausible explanations arising from the not entirely comparable circumstances. The Garland nomination occurs during the heat of the election year (starting almost coinciding with the first primaries) and runs all the way until a month afterwards. Plausibly, general political communication during this phase is especially likely to be aimed to be manipulative: for the short term goal of electing a President, towards an audience of people who are of varying degrees of engagement, swamped by many competing claims for attention. The Garland nomination may have begun and ended in futility: within days of the President's announcement, the Senate Majority Leader and the Chairman of the Judiciary Committee both announced their opposition to the confirmation, Majority Leader McConnell refusing to hold a vote, and Chairman Grassley refusing to even hold a hearing. The appropriateness of their refusal is not on point: without the agreement of the Senate leadership, the nomination was extraordinarily unlikely. The Garland nomination may have represented more grist for gratuitous opposition and symbolic support than the Gorsuch nomination. The Gorsuch nomination was strongly supported by the Republican Congressional leadership from the start, and thus may have necessitated a more reasoned opposition for any hope that he would fail to be confirmed to be remotely credible.

The Kavanaugh nomination was something different entirely. President Trump nominated Brett Kavanaugh, a well-known and acclaimed, if quite conservative, Federal judge to replace Anthony Kennedy, who had been a pivotal vote on important cases, most notably, to sustain legal abortions and the Affordable Care Act. And while Kennedy was far from a Supreme Court liberal in the mode of Ruth Bader Ginsburg, and voted with the conservative majority for most of his tenure, there was a heightened urgency (verging on panic) about the frailty of case law on these matters. Additionally, Kavanaugh had spoken publicly about the need for greater deference to the executive branch and expressed concern about the scope of the ongoing investigation by special investigator Robert Mueller into Russian interference with the 2016 Presidential election (and possible cooperation with the Trump administration). Until the last week of September, the critical issues in the Kavanaugh nomination concerned the likelihood that he would vote to overturn *Roe v Wade* or that he would accede to Trump administration demands to terminate the Mueller investigation. While enormously dispiriting towards supporters of abortion rights, with a Republican majority, the Kavanaugh nomination was likely headed for confirmation.

But then on 12 September 2018, Senator Dianne Feinstein shared with Senate Democrats the contents of a devastating letter by Professor Christine Blasey Ford reporting that Ford had been sexually assaulted by Brett Kavanaugh while the two were both in high school. On 13

September, Feinstein released a public statement that Feinstein had reported Ford's letter to the Justice Department. On 16 September, Prof. Ford identified herself to the *Washington Post* as the accuser, and provided vivid details. In the days that followed, two other women came forward with charges of sexually abusive behavior by Kavanaugh during his time in high school and at Yale.

Arguably, the conditions for the Kavanaugh nomination looked very much like those for the Gorsuch nomination until mid-September: although at a politically weaker point in his presidency and with attention centered on the upcoming midterm elections, the Kavanaugh nomination was nearly as assured as the Gorsuch nomination. As of mid-September, the possibility that key senators would vote against Kavanaugh's nomination became much more live.

The three nominees yield three quite different series due to starkly different political conditions, changes in the nature of the debates, and (for the Kavanaugh nomination) even in the data available. Further inquiry into the nature of *who* is sending the tweets can shed a little light on the streams.

5 Who (or what) are the tweeters?

While the presentation of the balance of deliberative and automatic tweets is helpful in and of itself, we are very much interested in the question of what kind of tweeter sends what kind of tweets. There are 221,455 unique users among the 52 million tweets in the Garland dataset.

Each time a user registers with Twitter, she or he (or it) supplies a Screen Name (which is associated with each tweet), and a brief self-description (which might change over time). In addition, each user U will acquire some important information as a consequence of interaction with Twitter: other users that follow U, as well as other users that U follows. These three bits of information might yield quite a bit of useful information. We further have the history, by user, of that users' tweets including the timestamp for each tweet.

The descriptions that the users provide permit a classification of who comprises the population of tweeters about the nominations. Here, we adopt multiple approaches. One is to employ a method similar to the sLDA discussed above to classify the tweets, but this time in the case of how the user identifiers her- or himself and as an unsupervised LDA. Even more simply, one might look at the bigrams of words used in the tweeters' self-descriptions. Some of the tweeters define themselves by family roles (e.g., "husband father/dad," "wife mother/mom,"), many by careers (e.g., "real estate," "freelance writer"), quite a number by interests ("social media," "political junkie," "social justice," "christian conservative").

The LDA led to further classifying information. Many tweeters also plainly identified their political leanings in hashtags like "#MAGA" or "#RESIST." Some included names of employers ("University of . . . ," or names of news organizations).

From this information, we were able to produce a number of occupational classifications, self-identified ideology, and self-identified strong position taking (aka "ideologues"). (Someone may be a liberal, an anti-Trump ideologue, or both). About 1.5% identified as liberal, .3% as anti-Trump, .8% as conservative, and .4% as pro-Trump.

Further, there is the problem of identifying "Bots," automata who may well come to dominate the flow of the quarter million tweeters during the nominations (and indeed did, but several criteria). There is a surprising degree of diversity in the intentional composition of the bots: some intend to influence the political conversation, some to distract from the conversation (King, Pan, and Roberts 2017), some to sell products (and simply hitch onto a trending twitter topic), and some to simply convey information (such as news organizations, or in the case of Garland especially, the local police or pet shelters). How might one identify whether a tweeter is a bot, or human?

The most obvious approach might simply be to look at the volume of tweets that any given tweeter might send, which ranged from 1 to 82,741. This approach to identifying bots is plausible, but imperfect. While there should be little doubt that the most active tweeters are probably automata and the least active tweeters probably humans, the boundary is a hazy one.

A second approach may be to look at the numbers of "friends" (tweeters followed by the tweeter) and "followers" (those tweeters who follow the tweeter). One might think that a typical human who has been on Twitter for a time would have a large number of both friends and followers, and that especially popular tweeters may have a large number of followers, or that avid social media junkies may follow a lot of others. In past research, it may well have been useful to employ a friends/followers ratio (bots have few other tweeters who follow them, while wanting to disseminate their messages to others). Except that in recent years, it has become evident that there are human tweeters who purchase followers[11], and even that networks of bots may follow one another. "Friends" and "followers" will be at best an imperfect indicator of the number of followers that one might have.

A third approach is to consider the entropy of the timeline of tweets, and is the one that we will depend upon in the present analysis. The concept of entropy denotes the amount of disorganization in a series, or unpredictability of the content. Claude Shannon (1948) famously identified a simple formula for the information entropy (H, read as eta) as

$$H(X) \;=\; -\sum_{i=1}^{n} P(x_i) \log_b P(x_i),$$

and where $b = 2$ is a typical choice for the base (to reflect the number of bits). (A pioneering paper on the use of Shannon Entropy is Chu, *et al.* 2012.) By virtue of the large number of tweets and a stochastic process interfering between when the tweet is sent by the user and the actual time recorded in the data, entropy scores range from 0 (for one-time tweeters) through about 8. Because some of the bots may be active for only a short period, we employ the monthly Average Entropy score instead of the Total Entropy.

The procedure has a robust pedigree, though we think our implementation is a flawed one,

[11] https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

where the flaw is due to a serious confound between the volume of tweets and an inevitable error that is introduced into the recording. Every time a tweet is sent, there is a stochastic lag time introduced before the tweet is recorded and disseminated by Twitter. Consequently, by measuring the timing down to the second, the distribution of the tweets will always include a random component. Since every tweet's time stamp is not just the time stamp sent, but the time stamp of the disseminated tweet, more tweets means more random noise. Consequently, our measure of $H()$ is strongly positively correlated simply with the volume of tweets.

(The pedigreed usage would note that humans tend to be more random than automata, which may release their tweets at the maximum rate per hour permitted by Twitter all at once. With comparable volumes of tweets, we should see *lower* $H()$ for automata than people; our usage here is exactly the reverse.)

We display these three measure of whether a tweeter is a bot or not in Figure 13. (The count of the total tweets is displayed in $\log_{10}$). It is quite apparent that two of the measures track each other quite closely: the Entropy score, and the $\log_{10}$ Count. The Count is quite a bit noisier than the Entropy measure, but the two are very highly correlated ($r = .91$). Because of the noise of the Counts as a measure of a underlying Bot trait, we will make extensive use of the Entropy score.

There are other measures of whether a tweeter is a bot or not (see Chu *et al.* 2012). One approach would be to examine the avatar for each tweeter for faces, although this is grossly imperfect. It is conceivable to use information about the social networks among the friends and followers, though that is more of a thought experiment than something we are considering. It is also conceivable to examine specific aspects about the timing of the tweets (e.g., regularity at the point when Twitter permits sending the maximum per hour), which is a promising line of exploration which we are currently exploring (Zhang and Paxson 2011).[12]

[12] Another approach is to use the "Botometer" rating service at Indiana University, which scrapes significant details about each tweeter's name to assess the probability that the tweeter is a bot under nine different variations. The process is computationally lengthy, and although complete for both the Garland and Gorsuch nomiations, the process for Kavanaugh (started as of Kavanaugh's confirmation on 4 October), still has not completed.

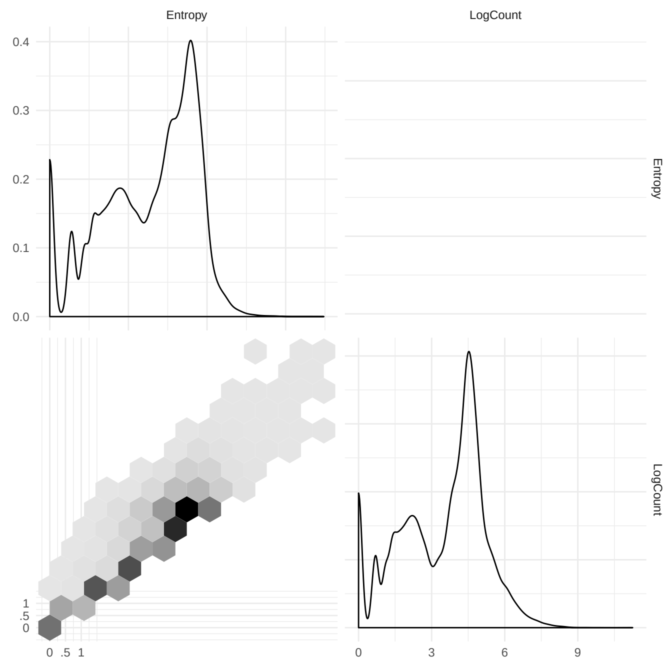

Figure 13: Entropy and the Volume of Counts measure the same thing

Lastly, we are able to compute an average sentiment score for each tweeter, and obtain a general measure of affect by the tweeter in tweets about the nominee simply by taking the average across each tweet of the sentiment scores computed by the sLDA and displayed in Figures 4–6. For those tweeters who were active in throughout the three nominations, we can also compute the difference.

6 Who sends what kind of tweet?

We then know quite a few things about the tweeters. but nearly all of these yielded red herrings or evidence that was inconsistent about the nominees. In analysis that is not reported here, we found:

- Self-identified ideologues were no more likely to be deliberative or automatic in their messaging than other tweeters;

- News organizations were more likely to be deliberative in the Garland nomination, but more likely to be automatic in the Gorsuch and Kavanaugh nominations. They were occasionally more likely to send more tweets that were neither deliberative nor automatic, but still on point (typically, statements of fact about procedures or meetings);
- Academics were just like other tweeters;
- Extreme bots were more likely to be automatic, but only for some months, than other users.

6.1 By Entropy and Ideology

We consider the behavior of the bots in the three nominations. We compute the similar first difference effects, except that this time we identify the Bots as those with Average Entropy greater or equal to 4 (at least at the level of an active political interest group, newsfeed, or commercial enterprise). We then combine the same information with the ideology of the tweeter as a way of distinguishing the "Liberal Bots" (e.g., NARAL's feed) from the "Conservative Bots" (e.g., Judicial Watch's feed). The three series appear in Figures 14, 15, and 16.

In general, Bots were no different from typical tweeters in the kinds of messages that they sent in the three nominations.The lines in both the automatic and deliberative tweets are about as flat as they can be (and where the perturbations are even smaller than those for the academics, the most inert group considered in this paper). While the result might be surprising to some, one should recall that the world of Bots is hugely diverse, with aims not just to persuade or manipulate an audience about the nominations, but also to do the same with regard to commercial products, as well as to provide general information. In the Garland series, this may have been particularly an issue since the volume of tweets still contained a lot of messages from politically irrelevant entities (e.g., pet shelters and nail salons in the Garland, Texas, metropolitan area), despite months of cleaning. But with the Gorsuch series, the problem of gathering irrelevant bots is less pronounced

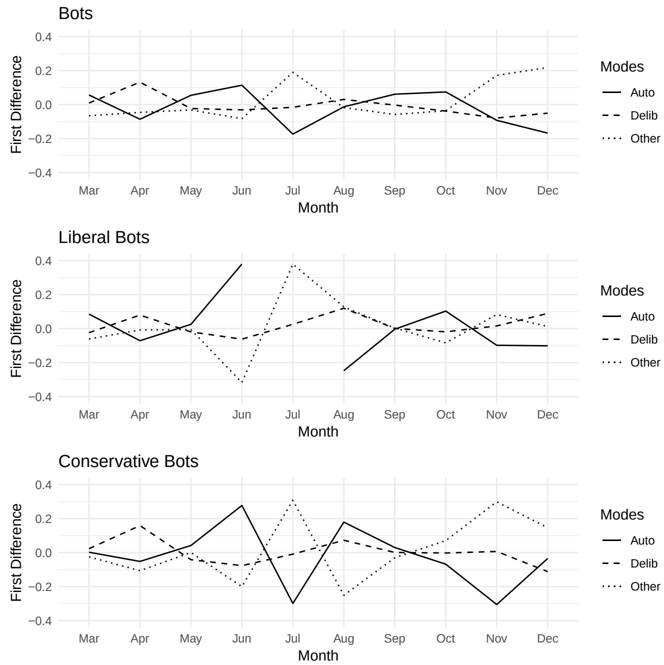

Figure 14: First differences in the probability of messages by type about Garland, Bots and Ideologues

given the uniqueness of the name. Further, some of the more active bots would have been news organizations and even municipal organizations, where the content of the tweet would have been strictly factual information about the state of the nomination (Logot and Diakopoulos 2015).

When we shift the picture to consider explicitly *political* bots, the matters are a bit more informative in the case of the Garland nomination. There is prior research that shows that bots can be *less* opinionated than humans (Dickerson, Kagan, Subrahamian 2014). Liberal bots were more likely to be manipulative in three months (May and June, by around 20%), and a lot less likely to be manipulative in August (also by about 20%). Liberal bots were no more likely than a typical tweeter to be deliberative, and even less so in June. Conservative bots were more likely to be manipulative in a couple of months (June at 7% and September at 10%), but a lot less manipulative than the typical tweeter in July (by 40%). Conservative bots were, like the liberal bots, no more deliberative than the typical tweeter with the small exception of June.

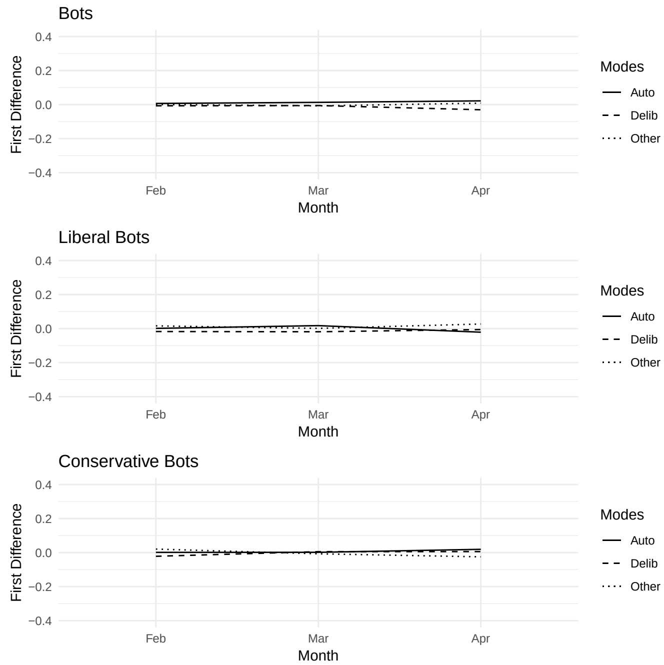

Figure 15: First differences in the probability of messages by type about Gorsuch, Bots and Ideologues

However, in the Gorsuch nomination the political bots are not really a factor. The Liberal and Conservative bots were no different from the typical user in the kinds of messages they were sending, whether those messages were predominantly deliberative ones (in February and March) or manipulative ones (in April).

The same results obtain for the bots, political or otherwise, in the Kavanaugh cycle. As with the Gorsuch plots, these are all flat lines.

6.1.1 By Strength of Sentiment

Are tweeters with strong sentiments about the nominees more likely to send automatic or deliberative messages? One hypothesis would be that since the tweeters are more emotionally invested in the nominees, they are more likely to be deliberative: the tweeter is more motivated, and if acting in a way that is concerned with the outcome, then we would expect to see more

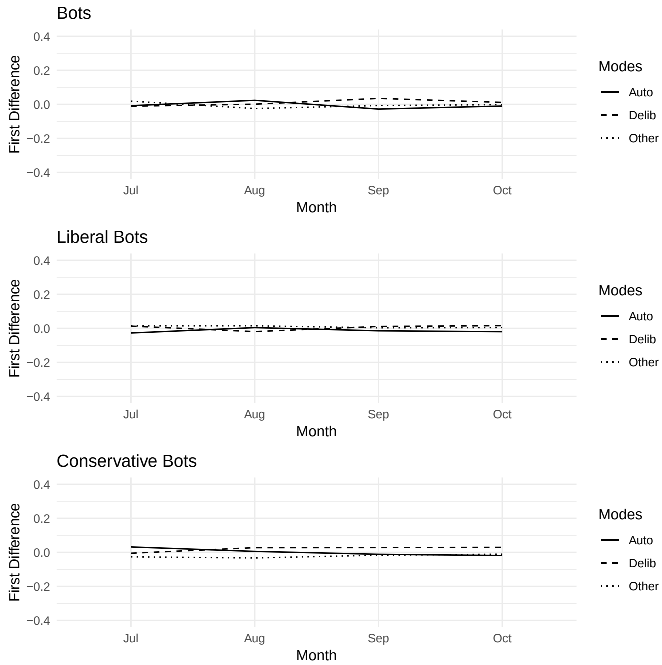

Figure 16: First differences in the probability of messages by type about Kavanaugh, Bots and Ideologues

deliberative messaging than automatic messages. If the outcome is a foregone conclusion, then deliberative messaging may be futile, and automatic messaging would dominate. There is some evidence among the nominees to support such a claim, but only in part.

As part of the first stages of this analysis, we computed the sentiment scores for each tweet initially as a simple word–lookup, and then as refined in the sLDA procedure itself (displayed in Figures 4 – 6). For each tweeter, we can compute the average sentiment score in the tweets about each nominee. In the analysis here, we use the absolute value of the average sentiment scores to capture strong feelings about the nominee.

The Garland nomination was futile from the moments that McConnell and Grassley announced that Republicans would not vote or deliberate on the nomination. If the hypothesis holds, we would expect more automatic messaging than deliberative ones, and Figure 17 shows as much. The three panels in these figures show the behavior of "medium" Bots (Entropy scores

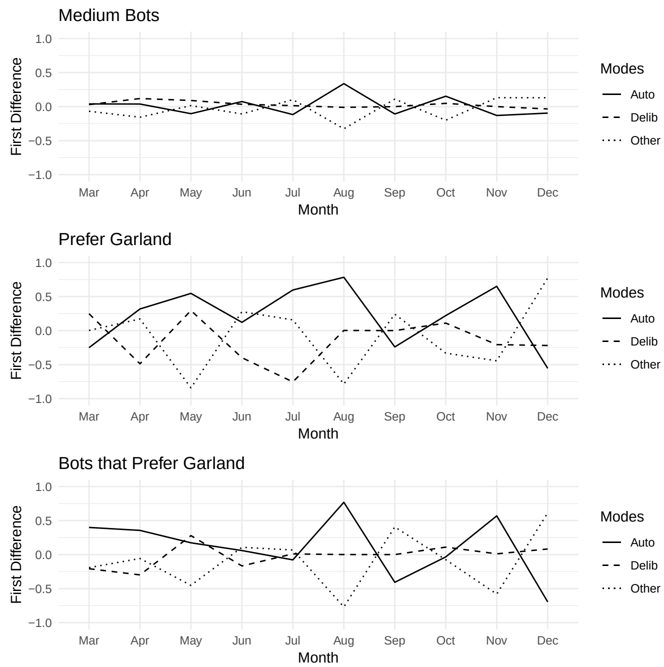

Figure 17: First differences in the probability of messages by type about Garland, Strength of Sentiment about Garland

over 4), the behavior of those with strong feelings about the nominee, and the interaction of the two. The top panel replicates the analysis of Bots in the prior pages, and shows that Bots were rarely different from other tweeters. The middle panel, however, shows that those tweeters with strong feelings about Garland were quite a bit more likely to be automatic than deliberative in seven of the months, while likely to be more deliberative than automatic in three of the months (May, September, and December). Bots that exhibited strong feelings about Garland were also more likely to send automatic than deliberative messages in all but four months (July, September, October, and December). In general, the Bots that exhibited strong feelings about Garland were more attenuated than the non-Bots.

Although the Gorsuch nomination was one that was in all likelihood inevitable, the Gorsuch tweets were also more likely to be automatic than deliberative, with the exception of Bots in the first month (Figure 18). The hypothesis would be that the tweets would be more automatic than deliberative, and for the last two months, the hypothesis holds: in both the panels for non-Bots

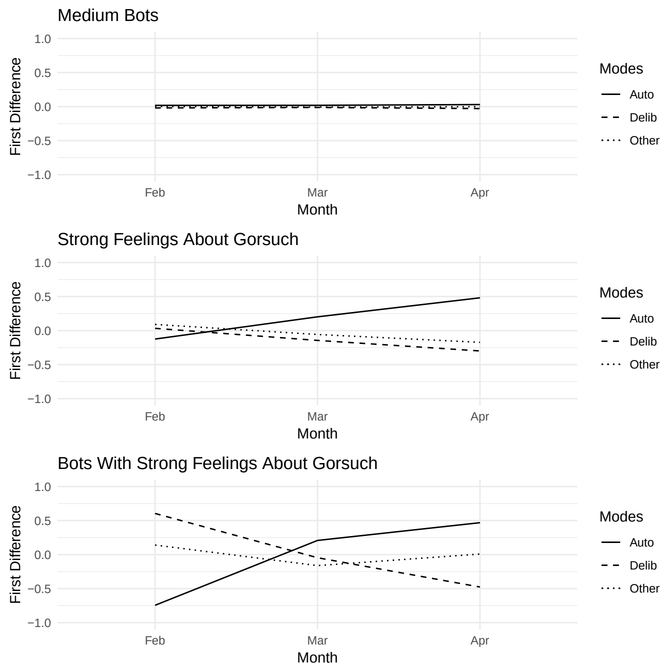

Figure 18: First differences in the probability of messages by type about Gorsuch, Strength of Sentiment about Gorsuch

and for Bots with strong feelings about Gorsuch, the first differences are positive and strong. But in the first month, Bots with strong feelings about Gorsuch (but not non-Bots) were much more deliberative.

The Kavanaugh nomination may have changed in inevitability in September. Although rued by liberal groups at the departure of Anthony Kennedy, none of the conditions about Brett Kavanaugh's nomination would have led one to expect anything other than confirmation: while there is clear evidence of controversial opinions by Kavanaugh in the past on abortion rights and the authority of the executive, Republicans maintained a (slim) majority in the Senate, the Bar Foundation found Kavanugh to be "highly qualified," and free (until mid-September) of anything other than minor scandals. Again, by the hypothesis, one would have expected the tweets to be more automatic than deliberative, and this holds for both July and August. But in mid-September, charges that as a high school student Kavanaugh had attempted to rape Professor Christine Blasey Ford began to circulate, and were eventually leaked to the press. There was a distinct possibility

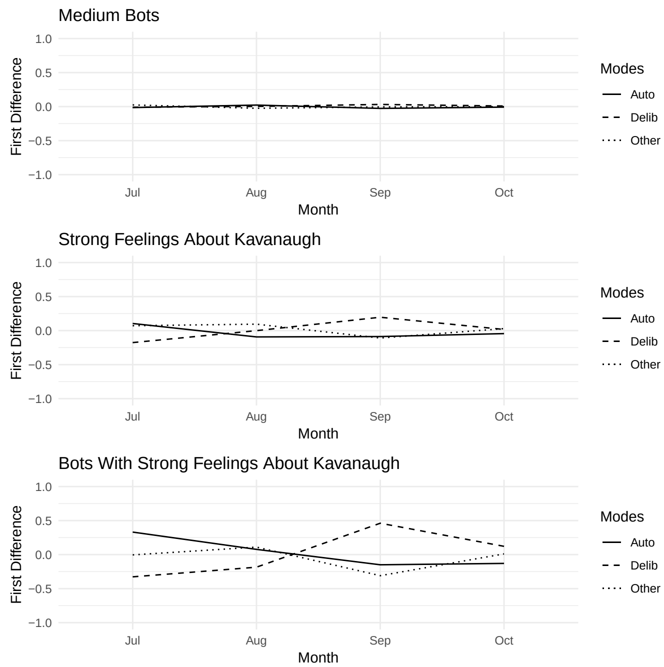

Figure 19: First differences in the probability of messages by type about Kavanaugh, Strength of Sentiment about Kavanaugh

that Kavanaugh would *not* be confirmed, even though Republicans held a majority. More of the tweets were deliberative in September and (the few days of) October than automatic (or "other").

The hypothesis that tweeters send messages that would have the effect of persuading people instead of manipulating people when the outcome is uncertain, but send manipulative messages when the outcome is known has provisional support. The one big piece of contrary evidence is from the very first month of the Gorsuch nomination among the Bots that exhibit strong feelings about Gorsuch, but otherwise, the claim that futility leads towards automatic messages appears to hold.

7 Conclusion

Our purpose in this paper is to examine the content of political messages over Twitter about two of the most salient (and stable) political questions in recent years, the nominations to the Supreme

Court of judges Merrick Garland, Neil Gorsuch, and Brett Kavanaugh. From extensive prior research in political psychology and political communication, the professions know that certain kinds of messages have consistent and demonstrable effects. But our argument in the present paper has been that we have not paid sufficient attention to what form these messages take. The form of the messages sent differs starkly between messages that have an enduring effect, and those that would only have a short-term effect. We cannot know the effect of political messages only by looking at the messages received.

In the course of this project, we believe we have made six separate methodological contributions:

1. We have explicate a method for handling a massive data stream of messages with respect to a specific political question. Twitter is a data source where the sheer volume of communications swamps by several orders of magnitude communications by newspapers and elected officials. According to Statista[13], there are about 68 million active Twitter users in the US alone, sending (world wide) more than 500 million tweets per day (or 200 billion tweets per year). Twitter is an unrivaled source of general conversation, and no small fraction of that conversation is political.

2. The sheer volume of information requires that we produce a method to winnow through the information. While there are excellent tools in both `R` and `Python` to systematically retrieve the tweets and to inquire about the tweeters themselves, actually making use of this information still demands that we cull. In one case, the culling required multiple levels of considerations, while in the others, the culling was far simpler. We have the code to systematically sort through this information.

3. We develop a rubric by which we might code the information into whether the message itself is a persuasive, manipulative, germane but neither persuasive nor manipulative, or irrelevant to the matter at hand. In this study, messages are persuasive if they are about the merits of the case, drawing on the extensive political science literature on the Senate confirmation process; messages are manipulative if they refer to extrinsic aspects of persuasion (such as the likability of supporting characters).

4. We argue for the application of a widely used statistical method (Supervised Latent Dirichlet Analysis) to classify the content of the messages into "buckets of words" according to the general sentiment of the messages themselves, and from the buckets, identify which rubric is most likely. The method then allows us to translate the association of each bucket with a rubric, the probability of the distribution of a message into each of the buckets, and ultimately the probability of belonging to a particular component of the classification by deliberative, automatic, other, or not applicable.

[13] `https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/`

5. We develop methods for assessing the self-descriptions of each of the message senders by classifying the self-descriptions into different occupations and political preferences.
6. We develop an approach based on conventional statistical analyses to assess the differences in the kinds of the messages sent by the tweeters based on what we know about the senders themselves.

The general character of the messages in the three nominations could not be more starkly different. The messages in the Garland nomination were overwhelmingly messages that were not about the merits of the case, but instead appeared to be more manipulative. The Garland nomination looks a lot more like the prior research on the messages in the *New York Times* and *Wall Street Journal* editorial pages (Brehm and Tutunik 2016), where we found that the manipulative (peripheral) messages dominated the persuasive (central) messages by about 19:1. The messages in the Gorsuch nomination flipped: far more of the messages were about the merits of the case and less about other routes to persuasion. The Kavanaugh nomination was something else entirely, beginning as a fierce debate about whether *Roe v Wade* would continue to be standing case law and the degree to which the President's decisions could be scrutinized by Congress and the courts, and then transforming very suddenly (and after the confirmation hearings had been concluded) into a bitter dispute about charges of attempted sexual violence by the nominee.

The largest difference between the three cases is the likelihood of an eventual confirmation. A popular Democrat in the final year of his presidency under divided government, and amid a heated campaign season with two historically different major nominees nominates a (slightly) left-of-center judge. An unpopular Republican in the first year of his presidency under unified government, and after a bitterly divided campaign of the historically different nominees nominates a (slightly) right-of-center judge. The first nomination was probably futile from the first few days after the declared opposition by the Senate leadership, yet remained fiercely fought through the end. In this first nomination, the messages tended to be manipulative (or likely to result in short term gains), but not about the merits of the case (or likely to result in long term gains). The second nomination was all but certain from the start with a well-qualified nominee supported by the Senate leadership, and while contested, was not fiercely fought. In this second nomination, the messages tended to be deliberative, and overwhelmingly so. The third nomination was close to certain (hanging only on the votes of Republican moderates who needed reassurances about *Roe,* in particular, but also deference to the executive), but became very much in doubt in the final month. The messages in the Kavanaugh nomination were initially more automatic than deliberative, but became overwhelmingly deliberative after the debate changes. The messengers appear to be choosing the character of persuading itself as the consequences of persuading matter more.

8 References

Abramowitz, Alan I. 1978. "The Impact of a Presidential Debate on Voter Rationality." *American Journal of Political Science* 22(3):680-90.

Ansolabahere, Steven and Shanto Iyengar. 1995. *Going Negative: How Political Advertisements Shrink and Polarize the Electorate.* New York: Free Press.

Ansolabahere, Steven, Shanto Iyengar, and A. Simon. 1999. "Replication Experiments Using Aggregate and Survey Data: The Case of Negative Advertising and Turnout." *American Political Science Review* 93(December):901–9.

Blei, David M. and John Lafferty. 2009. "Topic Models." In A. Srivistava and M. Sahami, eds., *Text Mining: Theory and Applications.* Taylor and Francis.

Blei, David M. and Jon D. McAuliffe. 2010. "Supervised Topic Models." In *arXiv:*1003.0783v1.

Brader, Ted. 2005. "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appeal to Emotions." *American Journal of Political Science* 49(April):388–405.

Brehm, John and Nicole Tutunik. 2016. "Persuasion in Policy Debate: Health Care." Unpublished manuscript.

Cameron, Charles M., Albert D. Cover, and Jeffrey A. Segal. 1990. "Senate Voting on Supreme Court Nominees: A Neoinstitutional Model." *American Political Science Review* 84(2):525–34.

Chu, Zi, Steven Gianvecchio, Haining Wang, and Sushil Jajodia. 2012. "Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?" *IEEE Transactions on Dependable and Secure Computing,* 9(10):1–14.

Cialdini, Robert B. 1984. *Influence.* New York: HarperCollins.

Dickerson, John P., Vadim Kagan, and V.S. Subrahamanian 2014. "Using Sentiment to Detect Bots on Twitter: Are Humans more Opinionated than Bots?" *Advances in Social Networks and Mining.* DOI:10.1109/ASONAM.2014.6921650.

Damore, D.F. 2002. "Candidate Strategy and the Decision to Go Negative." *Political Research Quarterly* 55(3):669–85.

Hillygus, D. Sunshine and Todd G. Shields. 2008. *The Persuadable Voter: Wedge Issues in Presidential Campaigns.* Princeton NJ, Princeton University Press.

Huber, Greg A. and Kevin Arceneaux. 2007. "Identifying the Persuasive Effects of Presidential Advertising." *American Journal of Political Science* (51(4):957–77.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. ?How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, not Engaged Argument.? *Ameri-*

can Political Science Review 111(3):484-501.

Lanoue, David J. 1992. "One That Made a Difference: Cognitive Consistency, Political Knowledge, and the 1980 Presidential Debate." *Public Opinion Quarterly* 56(2):168–94.

Lokot, Tetyana and Nicholas Diakopolos. 2015. "News Bots: Automating news and information dissemination on Twitter." *Digital Journalism* DOI:10.1080/21670811.2015.1081822.

Mansbridge, Jane J. 1983. *Beyond Adversary Democracy.* Chicago: University of Chicago Press.

Marcus, George E. and Michael MacKuen. 1993. "Anxiety, Enthusiasm, and the Vote." *American Political Science Review* 87(3):688–701.

Marcus, George E., W. Russell Neuman, and Michael MacKuen. 2000. *Affective intelligence and political judgment.* Chicago: University of Chicago Press.

McCullagh, P. and JA. Nelder. 1989. *Generalised Linear Models.* London: Chapman and Hall.

Mutz, Diana C., Richard A. Brody, and Paul M. Sniderman, Eds. 1996. *Political Persuasion and Attitude Change.* Michigan:University of Michigan Press.

Petty, Richard E. and John T. Cacioppo. 1986. *Communication and Persuasion: central and peripheral routes to attitude change.* New York: Springer–Verlag.

Shannon, Claude E. 1948. *The Bell Systems Technical Journal.* 27:279–423,623–656.

Song, Jonghuk, Sangho Lee, and Jong Kim. 2011. "Spam Filtering in Twitter using Sender–Receiver Relationship." Unpublished manuscript.

Sulfridge, Wayne. 1980. "Ideology as a Factor in Senate Consideration of Supreme Court Nominations." *The Journal of Politics* 42:560–7.

Yawn, Mike, Kevin Ellsworth, Bob Beatty, and Kim Fridkin Kahn. 1998. "How a Presidential Primary Debate Changed Attitudes of Audience Members." *Political Behavior* 20(2):155-81.

Wamg, Alex H. 2010. "Detecting Spam Bots in Online Social Networking Sites: A Machine–Learning Approach." In Sara Foresti and Sushil Jajodia (Eds.), *Data Applications Security and Privacy XXIV,* pp 335–43. Springer–Verlag.

Zhang, Chao Michael and Vern Paxson. 2011. "Detecting and Analyzing Automated Activity on Twitter." *Passive and Active Measurement,* p. 102–111.