Polarized, Together: Comparing Partisan Support for Trump's Tweets Using Survey and Platform-based Measures

Kenneth Joseph,[1] Briony Swire-Thompson,[2] Hannah Masuga,[3] Matthew A. Baum,[3] David Lazer[2]
[1]Computer Science and Engineering, University at Buffalo
338 Davis Hall, Buffalo NY, 14260
[2] Network Science Institute, Northeastern University
177 Huntington Ave., Boston, MA, 02115
[3] Kennedy School of Government, Harvard University
79 John F. Kennedy St, Cambridge, MA 02138
kjoseph@buffalo.edu, b.swire-thompson@northeatern.edu, hannah_masuga@student.hks.harvard.edu,
matthew_baum@hks.harvard.edu, d.lazer@northeastern.edu

Abstract

Using both survey- and platform-based measures of support, we study how polarization manifests for 4,313 of President Donald Trump's tweets since he was inaugurated in 2017. We find high levels of polarization in response to Trump's tweets. However, after controlling for mean differences, we surprisingly find a high degree of agreement across partisan lines across both survey and platform-based measures. This suggests that Republicans and Democrats, while disagreeing on an absolute level, tend to agree on the relative quality of Trump's tweets. We assess potential reasons for this, for example, by studying how support changes in response to tweets containing positive versus negative language. We also explore how Democrats and Republicans respond to tweets containing insults of individuals with particular socio-demographics, finding that Republican support decreases when Republicans, relative to Democrats, are insulted, and Democrats respond negatively to insults of women and members of the media.

Political polarization refers to the divergence of political ideologies between groups of citizens. In the United States, political polarization appears to have grown, as Republicans and Democrats increasingly disagree on a variety of issues (Doherty 2017). However, scholars disagree over the severity and real-world implications of this polarization (Fiorina and Abrams 2008). The current paper investigates the extent of political polarization in the United States in the context of public support for tweets sent by President Donald Trump, who, by most accounts, is a highly polarizing figure in American politics.

Specifically, the current study aims to investigate polarization in both "absolute" and "relative" terms. In absolute terms, Democrats are likely to show lower support than Republicans for any tweet from Trump. However, there could still be bipartisan support *relative* to a party baseline, or to another tweet. For example, do both Republicans and Democrats agree that a tweet about a U.S. Women's hockey team Olympic gold medal is better than a tweet insulting Hillary Clinton? Or do Trump's partisan insults attract enough support from his base to overwhelm general support for an American achievement? Three potential outcomes present themselves. The first is *hidden agreement* (i.e. a positive correlation between Democrat and Republican support across many tweets). For example, with hidden agreement, members of both parties would agree that gold medal tweet is better than the tweet insulting Hillary. The second is *true polarization* (i.e. a negative correlation), where polarization's hallmark of divergence is observed- where one group approves, the other disapproves and vice-versa. Finally, *ideological inconsistency* (i.e. no correlation) could occur, where there is no consistent pattern in relative support. In the present work, we analyze which of these three potential explanations best matches our data and then work to explain the correlation we observe.

By studying the American public's reaction to Trump's tweets, we are able to study patterns of absolute and relative agreement on thousands of individual data points, each varying in important and identifiable ways. We focus on Trump's tweets, as opposed to other politicians, because (1) they play a central role in today's political news cycle, (2) Trump is an extremely polarizing political figure (Swire et al. 2017), and (3) Trump's use of Twitter as the president of the United States is unprecedented. An American president has never before used a broadcast platform with such frequency or personal emphasis. According to Gallup[1], while only 8% of Americans actually follow Trump on Twitter, some 53% are regularly exposed to his tweets via other media platforms. Further, Trump's tweets have very real impacts in American society. One recent example is when the Dow Jones industrial average fell nearly 1,200 points in response to Trump's tweet about tariffs[2]. This suggests that individuals are not only being exposed to Trump's tweets, but that the tweets are impacting behavior in important ways. Given the nov-

[1]http://news.gallup.com/poll/234509/deconstructing-trump-twitter.aspx

[2]https://www.businessinsider.com.au/trump-tweet-dow-jones-stock-market-crash-economic-news-2018-2

elty of Trump's tweets and the extent to which they permeate American political discourse, a better understanding is necessary of Trump's daily musings, their resonance with the American public, and how they are perceived by members of both parties. Further, given the polarizing nature of Trump's tweets, agreement about particular tweets might signify important areas of unity across party lines.

In order to study public support for Trump's tweets, we must first ask a methodological question- namely, *how should we measure partisan support for a tweet?* One approach would be to use an "off-platform" measure - i.e., to survey offline individuals on their opinions of the tweet. A second approach would be to consider "on-platform" measures, such as the number of likes, retweets of and/or replies to the tweet.

Both survey and platform-based approaches have merits and drawbacks. Survey data often provides the most direct means of measuring public support. However, surveys are also often subject to small sample sizes, low response rates (Keeter et al. 2017) and the potential that self-report responses do not reflect respondents true opinions (Vaisey 2014). Twitter-based measures of public opinion, on the other hand, can alleviate issues of sample size, and because they measure actions, rather than opinions, can alleviate some forms of response bias by providing us with implicit measures of support. However, significant challenges exist for on-platform measures as well (Beauchamp 2017). For example, it is widely known that tweet metrics (e.g. number of retweets) can be disrupted by social bots (Ferrara et al. 2016). Further, actions on social media often carry social implications that may mediate behavior as individuals attempt to construct an idealized online persona (boyd, Golder, and Lotan 2010). We therefore compare both survey and platform measures for a comprehensive analysis of public support.

In the present work, we compare survey- and platform-based measures of public support for 4,313 of Trumps tweets posted between February 4th, 2017 and December 10th, 2018. For our survey-based measures, we take data from the YouGov TweetIndex (YouGov 2016). The TweetIndex provides survey-based responses from a group of opt-in participants on the YouGov survey platform. YouGov asks several hundred self-identifying Democrats and Republicans to rate each tweet on a five point Likert scale from Terrible to Great. Results are then made publicly available as distributions across political orientation. We aggregate these responses into a single measure per tweet, following the approach taken by YouGov.

For our platform-based measures, we use a dataset of over 1.8M Twitter users linked to American voter registration records. Because over 600K of these individuals are registered with either the Republican or Democratic political parties, we can readily analyze partisan support for each of Trump's tweets. Further, by linking accounts to voter records with high fidelity, we reduce the risk that our measures of support are tainted by bots or non-Americans. Note that we exclude non-Americans in this study because of a focus specifically on Americans' opinions. Using this dataset of Twitter users, we develop a platform-based measure of both Democrat and Republican support for each of Trump's tweets. As we will show, it is important to incorporate both retweets and replies into a metric of support, rather than relying solely on one or the other.

The first contribution of this work therefore concerns the methodological question of how to measure partisan support for tweets. We show that, provided we construct a suitable metric, we can weakly approximate survey-based measures with platform-based measures. However, we also identify important differences across measures, suggesting a need for caution in interpreting only platform-based measures. The second contribution of our work is an analysis of partisan support for Trumps tweets. Our main findings are four-fold:

- As expected, we observe that there are large differences across party lines in support for Trumps tweets. This holds across both types of support measures we analyze.
- However *after controlling for mean partisan differences, Republicans and Democrats show significant agreement on both survey- and platform-based measures of support.*
- We observe that partisan agreement can be partially explained in part by the fact that both Republicans and Democrats dislike when Trump insults others, uses negative sentiment, tweets content that is demonstrably false, and like when he uses language supporting the military/first responders, or tweets his condolences. Republicans are, however, much less likely to shift away from a positive view of Trump, regardless of tweet content.
- Analyzing tweets containing personal insults, we find across both support measures that Republicans showed higher levels of support when the target of the insult was a Democrat, relative to a Republican, and that Democrats showed lower levels of support in response to insults of women or members of the media.

Related Work

We discuss relevant work related to the study of political polarization and measurement of politically-relevant content on social media.

Political Polarization

Surveys report that Americans are increasingly polarized in their ideological preferences (Doherty 2017), and behaviors on social media also reflect strong partisan divides (DellaPosta, Shi, and Macy 2015). It is important to realize, however, that various forces are at work in shaping observed polarized ideological preferences. For example, political scientists have argued that in recent years we are observing the effects of *sorting*, where people with liberal ideology have better organized themselves into the Democratic party and those with conservative ideology increasingly identify as Republican (Mason 2015; Levendusky 2009). This is distinct from the process of political polarization as typically conceptualized, wherein individuals' political views shift to be more extreme.

Perhaps more importantly, scholars have argued that observed ideological polarization may be driven by a strong attachment to one's partisan identity rather than a coherent set of beliefs. Americans demonstrate less ideological separation than they believe themselves to have (Levendusky and Malhotra 2016), and relatively few show consistent conservative or liberal ideological standpoints across all issues (Converse 1964). Scholars have instead shown that polarization, at least in survey responses, may be a function of post-hoc rationalization or motivated reasoning (Kunda 1990; Lord, Ross, and Lepper 1979)—where individuals attempt to answer “like a Republican” rather than in a way that aligns with their personal beliefs.

This argument therefore suggests that the driver of partisan divides in empirical statements of ideology might be due not to differences in ideologies, but rather to social identity-based factors; specifically, increasing attachment to one’s own party and animosity towards members of the other political party (Iyengar and Westwood 2015). Barber and Pope (2018) explore this hypothesis in the context of conservatively and liberally aligned statements made by Trump. They find that party loyalty, especially for Republicans, is a much stronger driver of survey responses than any consistent ideological viewpoint.

It is therefore possible that while Republicans and Democrats are polarized in absolute support for Trump’s tweets, they may show relative agreement due to a lack of coherent ideological beliefs. The current study thus complements the work of Barber and Pope (2018) by asking similar questions about tweets, rather than about content with particular ideological meanings. The benefit of doing so is that we can consider polarization across thousands of observations, rather than a limited number of ideologically aligned views, making a more general claim about correlations of support across party lines. Additionally, we expose a variety of other factors unrelated to ideology that impact support.

Measurement of politically-relevant content

In order to further investigate patterns of political polarization, it is necessary to engage with the broader literature on using social media to study public opinion. Scholars have studied the connection between social media data and public opinion as expressed in various forms, from polling data (O’Connor et al. 2010; Beauchamp 2017) to elections (Hobbs et al. 2017) to movie box office sales (Asur and Huberman 2010). Perhaps most relevantly, Beauchamp (2017) critically reviewed existing work on poll prediction and presented a novel penalized-regression based model and evaluation framework for poll prediction models. We build on Beauchamp’s (2017) work but focus on a task that differs in two ways. First, prior work largely focuses on prediction. While an important future goal, the present work focuses on explaining existing data rather than predicting future events. Second, in contrast to focusing on a macrosocial signal, such as polling data, for which tweets are aggregated across many users to make predictions about one number (e.g. presidential approval ratings), we essentially focus on the opposite problem - how do we measure broad support for a single tweet?

With respect to the construction of a platform-based measure, previous literature has suggested that retweets could potentially be used as an indicator of support, where replies indicate disapproval. A long line of work has considered the meaning of retweets (Macskassy and Michelson 2011; boyd, Golder, and Lotan 2010; Metaxas et al. 2015; Guerrero-Solé and Lopez-Gonzalez 2017). A recent metareview and survey analysis suggests that *official* retweets—that is, retweets that use the retweet button on Twitter—are largely signals of support. Survey respondents stated that they tended to retweet content they found credible, trustworthy, interesting, agreeable, or entertaining from accounts they believed were credible and trustworthy (Metaxas et al. 2015). However, other work has suggested that retweets may not always indicate support (Macskassy and Michelson 2011), particularly when they are expressed via unofficial means (e.g., via adding the text “RT” to beginning of the tweet (Azman, Millard, and Weal 2012)). Even via the official mechanism, at least in highly polarized domains like sports and politics, between-group retweet rates can actually exceed within-group rates. This difference can be partially explained, however, by the time between when the tweet was sent and when it was retweeted (Guerra et al. 2017). We find similar patterns in Democrat retweeting behavior with respect to Trump’s tweets.

Less well-studied is whether or not we can broadly characterize other forms of interactions with tweets besides retweeting—like quoting, replying, or liking—as mechanisms of support of disapproval. Garimella, Weber, and De Choudhury (2016) study quote tweet and reply tweet behaviors, finding that typical users generally employ quotes to publicly reply to, to express an opinion of, or to forward content from an original tweet. Additionally, they found that 66% of replies and 58% of quotes were insults, suggesting a broad use of replies as negative responses to content. Consequently, it may be reasonable to broadly employ replies as a rough indicator of disapproval, while quote tweets are perhaps used in too many different ways to generalize their meaning.

These academic findings are consistent with a popular view that retweets generally indicate support, and that replies generally indicate disapproval for a tweet. For example, Roeder, Mehta, and Wezerek (2017) detail “The Ratio”, a comparison of a tweet’s likes, retweets and replies. They use this to analyze support for tweets of various U.S. politicians, including Trump. Their work falls within a broader set of non-academic studies of Trumps tweets. Robinson (2016), for example, shows that tweets from Trump’s account by an Android phone, presumably from Trump himself, differed in various ways from tweets likely sent by staffers using iPhones. In an analysis we return to below, Lee and Quealy (2016) identify targets of insults from Trump. These analyses suggest important ways in which Trumps tweets can be dissected and further understood, although no work we are aware of to date has rigorously considered how either survey- or platform-based measures of support might be explained by features of individual tweets.

Finally, it is worth noting that scholars have also investigated how politicians express themselves online. Stanyer (2008) reviews related literature and provide a cross-cultural analysis of politicians’ online self-presentation in the U.S. and U.K. McGregor, Lawrence, and Cardona (2017) further

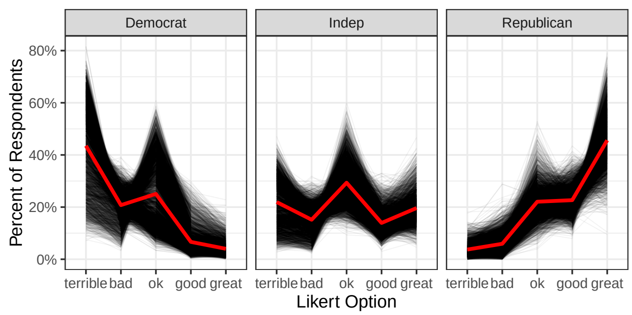

Figure 1: Responses on the Likert scale for the YouGov TweetIndex survey data. Each black line represents a single tweet. For each tweet and Likert condition, the y-axis represents the percentage of respondents who gave that answer on the Likert scale. The red line represents the average across all tweets. We, like YouGov, split the data across Republicans, Independents and Democrats.

finds that identity creation strategies of politicians are gendered, with male and female politicians adopting different strategies in the face of expected gender stereotypes. Our work sheds further light on the ways in which Trump and his staff present his online persona and how the American people receive these decisions.

Data and Methods

We leverage data from a variety of sources. Here we detail how we a) used YouGov data to compute a survey-based measure of support for Democrats and Republicans for each tweet, b) used our panel of Twitter users linked to voter registration records to compute a corresponding platform-based measure, and c) how we developed variables used to explore patterns in these four support measures. All analyses and, where possible, all data used is available as part of a Github repository at https://github.com/kennyjoseph/trump_tweets_icwsm.

Survey-based support measures

On December 10th, 2018 we extracted survey responses to 4,403 of Trump's tweets from the YouGov Tweet Index. This set contains every tweet Trump sent between February 4th, 2017, shortly after his inauguration, and December 10th, 2018. YouGov is a well-respected polling firm whose data has been widely used in academic studies, particularly in political science (Twyman 2008; Ansolabehere and Schaffner 2014). We removed 12 of these tweets that were part of multi-tweet chains[3], 66 tweets sent between December 1st and December 10th for which we did not have data from our Twitter panel, and 12 deleted tweets. This left us with a final set of 4,313 tweets. For each tweet, YouGov asked a sample of American adults to rate the tweet on a five point Likert Scale, with the options "Great", "Good", "OK", "Bad" and "Terrible". Figure 1 shows that the modal score for Republicans was "Great", for Democrats was "Terrible", and for Independents was "OK". In the figure, each grey line represents a single tweet, and the red line the average percentage of respondents across all tweets that gave a particular answer.

In order to compare with platform-based measures, we characterize each tweet by a single value. YouGov chose to do this by aggregating responses across several hundred Democrats, Republicans and Independents by taking the average numerical value assuming "Great" is +2, "Good" is +1, etc., and multiplying by 100. We assessed various other means of aggregating scores (e.g. exponential weighting schemes), all of which were highly correlated with the scoring mechanism devised by YouGov. Consequently, we retain this scoring measure. Additionally, given our interest specifically in partisan views, we do not consider a measure for Independents.

Platform-based support measures

We have developed a panel of approximately 1.8M Twitter users linked to voter registration records, using methods similar to those in prior work (Barberá 2016). At a high level, we begin with a large set of both Twitter accounts and voter registration records, and then match accounts to voter registration records if a) they have the same name and location, b) no other Twitter account exists with that name that has no identifiable location attached to it, and c) the name and Twitter account are both unique within a given U.S. city or, if a city could not be discerned, a state.

More specifically, we begin by collecting a large sample of Twitter users (approximately 406M) who sent one or more tweets that appeared in the Twitter Decahose from January 2014 to August 2016. In March of 2017, we used the Twitter API to select from this set the approximately 322M accounts that were still active. Using time zone and language information from their profiles, we then remove accounts that are clearly located outside the United States. Using the name field and screen name, we extract a set of "name words" from each profile to use for matching names. We further exclude profiles having fewer than two name words, leaving a total of around 237M Twitter accounts to match voters against. Having curated this set of Twitter users, we next turn to voter registration records. We obtained voter data from TargetSmart, a non-partisan source of voter registration data. TargetSmart provided a comprehensive set of name and address records for U.S. individuals, including those not registered to vote.

We perform location extraction from Twitter profiles using a set of rules on the location field of the profile, where we attempt to extract a city and/or state level location. For instance, from profiles listing "Buffalo, NY" or "New York", we would infer locations of "Buffalo, New York (State)" and "New York (State)," respectively. Finally, we then attempt to match these to the respective fields in the voter registration data. For more details on the matching process, we refer the reader to Grinberg et al. 2019, where the same matching approach was used on a smaller sample of voter registration records. Manual evaluation of our matching pipeline

[3]it was not clear whether or not respondents saw all tweets on the chain

indicated that our conservative approach yields precision on the order of 90% (Grinberg et al. 2019). Additionally, we found that while the population we focus on - people who provide their real names and locations on Twitter- is no doubt skewed, demographics of the matched population match well with survey data on the broader population on Twitter across a variety of demographic characteristics, including age, gender and race.[4]

Using this curated panel of Twitter users, we can be reasonable confident that we are studying the behavior of real Americans. For a portion of the panel (36.3%), we are further able to determine a political affiliation where panel members are registered with either the Republican or Democratic Party. In total, we have 246,509 registered Republicans, and 398,945 registered Democrats. For these individuals, we identify each retweet of and reply to Donald Trump's tweets and aggregate these metrics across Republican and Democratic registered voters. Given the novelty of quote tweets and the variation in their use (Garimella, Weber, and De Choudhury 2016), we do not consider them here. In total, we find that at least one Democrat or Republican interacted with 21,617 of Trump's tweets, including all tweets within the span of the YouGov data as well as, obviously, many tweets prior to YouGov's data collection.

Figure 2 displays four scatterplots comparing retweeting and replying-to behaviors of Democrats and Republicans, showing that Democrats reply more to Trump, while Republicans retweet Trump more. This observation is in line with prior work discussed above, where retweets are a signal of support and replies a (weaker) signal of discontent. The exception is tweets sent before Trump was inaugurated (points colored grey in the plot), which are often retweeted more by Democrats than Republicans and replied to more often by Republicans than Democrats. A straightforward explanation exists for this, matching well with insights from prior work (Guerra et al. 2017): Democrats often retweeted old tweets of Trump that contradicted his current actions, with Republicans replying to combat this behavior. For example, the most retweeted Trump tweet by Democrats, retweeted by .4% of those in our sample, was "Are you allowed to impeach a president for gross incompetence?", a tweet sent by Trump in 2014.

Figure 2 suggests that both retweets and replies provide useful information for a measure of support. However, looking at the text of the most replied to and/or retweeted tweets within the YouGov sample suggests the picture is slightly more complicated. For Democrats, the three most replied-to tweets include two anti-global warming tweets and one about the Muslim travel ban, tweets we might expect Democrats to disapprove of. For Republicans, however, the three most replied-to tweets included a tweet bashing CNN, which we would *not* expect Republicans to disapprove of.

[4]The use of this panel for this study has been approved by Northeastern University's Institutional Review Board (IRB). Further, note that we link Twitter users to voter records only in cases where the Twitter users have provided their full (and real) name and location. This falls within the guidelines of Twitter's Terms of Service, which stipulate that linking Twitter data with offline data is acceptable under reasonable expectations of privacy.

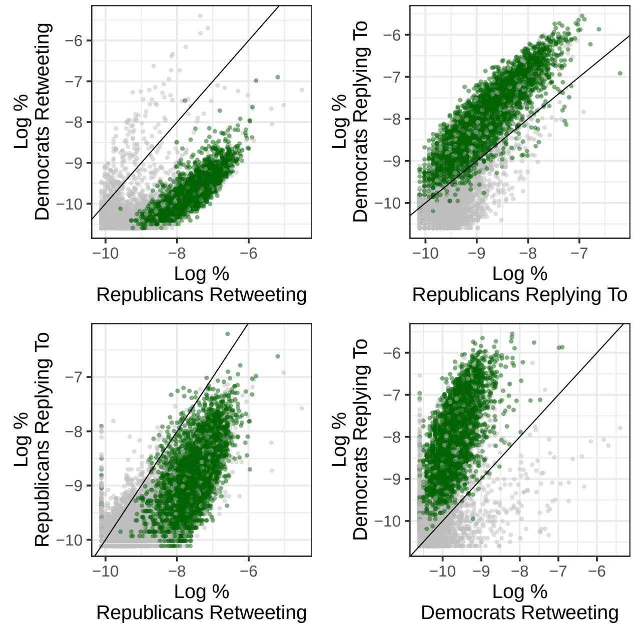

Figure 2: Four scatterplots, presenting the logged percentage of Republicans or Democrats replying to or retweeting a tweet. A small constant is added to avoid undefined values. Each dot is a tweet sent by President Trump. Points in scatterplots are colored by whether or not they are in the YouGov data - green points are included in the YouGov data, i.e. were sent after Trump's inauguration, grey points are not (were sent pre-inaguration). The black diagonal line in each plot is a line with a slope of 1 and an intercept of 0 (i.e. the line $y = x$)

Moreover, this same tweet was the most retweeted by Republicans *and* Democrats. Reply and retweet counts therefore contain sometimes conflicting information and, as prior work has shown, varying motivations. We therefore combine these two signals to attempt to smooth over these variations. An additional argument in support of a combined measure is that on their own, (logged) retweet and reply counts are only weakly correlated, or entirely uncorrelated, with our survey-based metrics. Retweet counts, which we expect to signal positive support, are actually *negatively* correlated with the survey-based measure for Democrats (-.26) and only weakly correlated for Republicans (.03). Correlation between survey-based measures and reply counts are higher (-.56 and -.32 for Democrats and Republicans, respectively)[5].

However, we can construct a simple metric using both retweets and replies that yields an even stronger correlation with survey-based measures. Specifically, we can construct a platform-based measure of support as the log-odds of a Democrat (Republican) retweeting as opposed to replying to a given tweet. Mathematically, this means that our platform-based measure of support for, e.g., Democrats, for

[5]Note that because replies are expected to generally be negative, we would expect such negative correlations

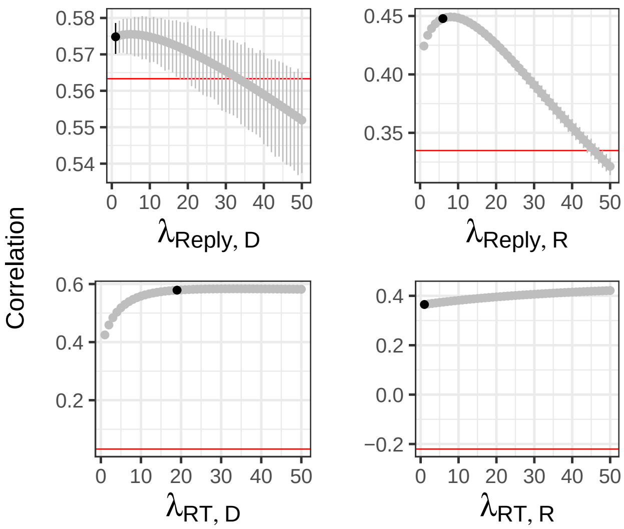

Figure 3: Results for simulation studies on smoothing parameters in Equation 1. Graphs provide 95% (bootstrapped) confidence intervals (CI) for correlation with the survey-based measure (y-axis) as the parameter value varies (x-axis). For example, the upper-left subplot shows how correlation between Democrat platform- and survey-based support measures varies as we change the smoothing parameter for Democrat replies, $\lambda_{Reply,D}$. CIs represent variance over different values of $\lambda_{RT,D}$. Red lines represent maximum correlations with the survey measure across potential smoothing values using only logged counts of replies or retweets (e.g. in the upper left, using only replies for Democrats)

a given tweet, t, is calculated as follows, where $\#RT_{t,D}$ is the number of Democrats who retweeted t, $\#ReplyTo_{t,D}$ is the number of Democrats who replied to t, and smoothing terms $\lambda_{RT,D}$ and $\lambda_{Reply,D}$ are added to ensure non-zero values:

$$\log \frac{\#RT_{t,D} + \lambda_{RT,D}}{\#ReplyTo_{t,D} + \lambda_{Reply,D}} \tag{1}$$

One final question is how best to determine appropriate values for these smoothing terms. As Monroe, Colaresi, and Quinn (2008) has argued, this decision is important especially when signal from a given data point is limited (e.g. there are few retweets or replies for a given tweet). We choose ideal values of $\lambda_{RT,D}, \lambda_{RT,R}, \lambda_{Reply,D}, \lambda_{Reply,R}$ using a simulation-based approach where we identify parameter values that maximize the correlation of the platform-based measures with the survey-based measures. We explain our simulation procedure for $\lambda_{RT,D}$ only and note that others are analogous. We first compute correlations for our platform-based measure with the survey-based measure varying both $\lambda_{RT,D}$ and $\lambda_{Reply,D}$ from 1 to 50. For each value of $\lambda_{RT,D}$, we then construct a one standard deviation confidence interval of this correlation across the different values of $\lambda_{Reply,D}$. Finally, drawing from approaches to regularized regression (Friedman, Hastie, and Tibshirani 2009), we select the value for $\lambda_{RT,D}$ that is within one standard deviation of the value of $\lambda_{RT,D}$ having the maximum average correlation with the survey measure.

Figure 3 provides results from this simulation study for each of the four parameters. We also vary a smoothing parameter for raw retweets and replies, and show the best obtained correlation as a red line on the corresponding plot[6]. So, for example, the upper left figure presents a red line for the maximum correlation between logged counts of Democrat replies and the Democrat survey-based measure over all possible values of a smoothing parameter. Figure 3 shows that in all cases—even when a naive smoothing parameter of 1 is selected for all 4 λ values—our platform-based support measure shows significantly higher correlations with the survey-based measure than using either retweets or replies on their own. Final values used were $\lambda_{RT,D} = 19, \lambda_{RT,R} = 1, \lambda_{Reply,D} = 1, \lambda_{Reply,R} = 6$.

Explanatory Analyses

Having developed measures of support, we now turn to how we explain patterns in them. We carry out two different explanatory analyses. First, we study patterns across the entire dataset, using both theoretically and practically-informed features. Second, we subset our analyses to only those tweets containing an insult of a person, and study how the various measures of support respond to insults of individuals with different sociodemographic characteristics.

Analysis of all tweets We leverage two external sources of data to help explain how the different support metrics respond to particular tweets. First, we leverage hand-coded data developed by *New York Times* journalists on insults made by Trump since he declared his candidacy (Lee and Quealy 2016). Importantly, these tweets are *insults*, and are not necessarily within the related-but-distinct domain of *hateful* or offensive content (Davidson et al. 2017). For example, the following tweet contains an insult directed towards Senate Republicans and Democrats, but is unlikely to be considered hateful or even offensive: "3 Republicans and 48 Democrats let the American people down. As I said from the beginning, let ObamaCare implode, then deal. Watch!" Second, we leverage hand-coded data developed by the *Washington Post* on the veracity of the information contained in Trump's tweets, starting after his inauguration and continuing to the present.

Using these two datasets, we construct two variables per tweet, one for whether or not a tweet contains an insult and a second for whether or not it contains a falsehood. We also construct three additional independent variables per tweet based on the content of the tweet itself. First, to assess how bipartisan agreement varies across the sentiment of Trump's tweets, we leverage the Twitter sentiment analysis tool VADER (Hutto and Gilbert 2014). While VADER reports a continuous score on the [-1,1] interval, score distributions were heavily peaked around -1, 0, and 1. We therefore use a tertiary variable that identifies whether or not a

[6]For replies, we depict the absolute value of the correlation

tweet was negative [-1,-.1), neutral [-.1,.1], or positive (.1,1]. Note that both conceptually and empirically, insults and sentiment are distinct.[7]

Second, hypothesizing that individuals on Twitter might be less likely than survey participants to watch or read linked content, we constructed a variable for each tweet based on whether or not it contains an external link. Finally, in initial analyses of the data we observed that a considerable number of Trump's tweets contained rhetoric about either supporting the military and law enforcement, or condolences after loss or tragedy. Such tweets, we found, seemed to show strong positive support across all four metrics. Consequently, we construct a final variable using a regular expression that identifies whether or not the tweet contains a word, stem, or phrase relating to these topics.[8]. We refer to this as the "support or condolence" variable.

We run linear regression models to identify how these factors are associated with different levels of support for Trumps tweets. To do so, we fit a single regression model across all four metrics. To include both survey-based and platform-based metrics in the same model, we center and scale scores for each type of measure (i.e. we center and scale twice, once for the two survey measures, and once for the two platform measures). We include each of the five independent variables, which we interact with two other variables: one for measure type (Survey or Platform) and one for party (Democrat or Republican). We also include main effects for measure and party type.[9].

Analysis of personal insult tweets In addition to identifying factors explaining variance across all tweets, we also focus specifically on cases where Trump insults individuals to assess how different measures may respond to insults towards different kinds of people. Trump's personal insults have been the source of significant controversy[10]. Furthermore, it seems that insults are often aimed at, or more critical of, individuals of a particular gender, race or political affiliation. We therefore wished to further explore partisan responses to these personal insults.

To identify personal insults, we extract the 130 people out of 550 total "people, places and things" that Lee and Quealy (2016) identify as being insulted by Trump.[11] Collectively, Trump's tweets insult these individuals 494 times in our dataset. We then, using information extracted from Wikipedia infoboxes, characterize each individual according to their gender (male or female), race (White, Black, Asian or Middle Eastern), political affiliation (Democrat, Republican, or Unspecified/Unclear), and primary occupation (member of the Media, Politician, Non-elected Government Official, Foreign Leader or Other/Unspecified).

We again run simple linear regression models to identify how these factors are associated with different levels of support for Trump's tweets across all four of our support metrics. We here treat each measure individually, scaling and centering all four measures independently.

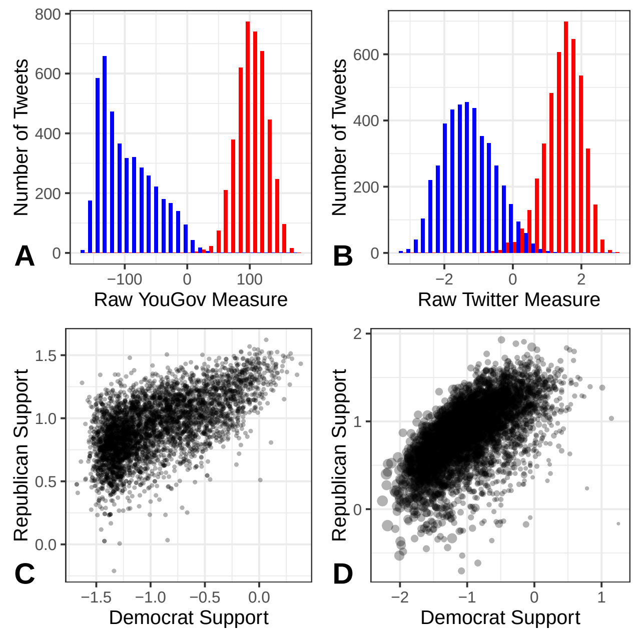

Figure 4: A) Histogram of survey-based measures for the 4,313 Trump tweets we study for Democrats (blue) and Republicans (red). B) The same information, except for the platform-based measures. C) Each point represents the value of the Republican support measure (y-axis) and Democratic support measure (x-axis) for the survey data for a single tweet. D) The same information, except for the platform-based measure. In D), points are sized by the number of individuals who interacted with the tweet.

Results

We first provide results summarizing survey and platform-based measures and correlations across them. We then turn to our two explanatory analyses.

Patterns in Support Measures

Figure 4a) and b) shows that both survey- (Figure 4a) and platform-based (Figure 4b) support measures are heavily polarized. Out of a total possible score in the range of [-200,200], the median survey-based measure for Democrats was -103 and for Republicans was +100. For platform-based metrics, the median was 1.46 for Republicans, meaning Republicans were roughly four times more likely to retweet a tweet than to reply to it, whereas for Democrats it was -1.27, meaning Democrats were roughly three and a half

[7]Insults, as a form of stance, have long been shown to be distinct from sentiment (Johnson and Goldwasser 2016). Indeed, in our data, 41% of tweets with insults have a positive sentiment.

[8]Words/stems/phrases used: *law enforcement, safe, hero,first responder, disaster, congrat, victim, tragic, bless, storm, evacua, serv, pray, hurricane, symp, happy, condol, brave*

[9]Note that even though we scale and center scores for each support measure, we can still estimate an intercept because it takes into account other variables in the model

[10]https://www.cnn.com/2018/08/18/politics/who-trump-attacks-insults-on-twitter/index.html

[11]All other insults listed were either to non-persons or were contained in tweets not in our dataset

times more likely to reply than to retweet. Note that because $\lambda_{RT,D} > \lambda_{Reply,D}$ and $\lambda_{RT,R} < \lambda_{Reply,D}$, these numbers are in fact slightly more conservative than a raw estimate would provide.

One additional point suggested in Figure 4a) and b) is that Republican measures are more concentrated than Democrat measures. Indeed, standard deviations of Republican scores are 75% higher for the survey measure and 39% higher for the platform measure. This suggests that Republicans, according to both measures, are less willing to shift their view of Trump based on content of the tweet itself. Even with this reticence by Republicans to respond differentially to particular tweets, however, Figures 4c) and d) shows that both support measures are positively correlated across partisan lines: .62 for the survey measure and .59 for the platform. While different on their overall support towards Trumps tweets, Republicans and Democrats demonstrate moderate agreement over which tweets are relatively better and which are relatively worse.

The question arises as to whether this hidden agreement emerges for other politicians as well. While we have no survey-based support measure for other politicians, the correlation between our two measures, even when smoothing parameters are set naively (and low), suggests that we can have some confidence in assessing results from only our platform-based metric. Using the same approach as described above, we construct the identical platform measures of support for the Twitter accounts of five politicians, three Democrats—Hillary Clinton, Barack Obama, and Nancy Pelosi— and two Republicans, Mitch McConell and Mike Pence. For each, we set $\lambda_{RT,D} = \lambda_{RT,R} = 3$, and $\lambda_{Reply,D} = \lambda_{Reply,R} = 1$, reflecting the fact that in general, retweets are more prevalent than replies, and remove tweets with less than ten total interactions, the minimum observed in our dataset of Trump tweets.

For all politicians, we find a significant and positive correlation between Democratic and Republican platform-based metrics. We also find that for all Democratic politicians, the median Democratic platform-based metric was higher than the Republican metric, and for Republican politicians, the Republican metric was higher. However, in several cases, members of the opposite party were still slightly more likely to retweet than to reply to a particular politician, and correlations between scores were lower than for Trump, ranging from .23 (Pelosi) to .40 (Obama). While these results support our findings of polarized support and hidden agreement, they also suggest that responses to Trump are potentially unique in light of his particular use of the platform.

Explaining Partisan (Dis)agreement

Of the 4,313 tweets we study, 28% contain an insult, 22% contain a false statement, 16% contain language related to support or condolence, 29% contain a URL, and 30%, 12% and 58% have negative, neutral and positive sentiment, respectively. Figure 5 shows univariate patterns in how each support measure responded to each of the independent variables we consider. Two broad patterns are apparent.

First, with one exception all five independent variables have the same directional effect on each of the four support

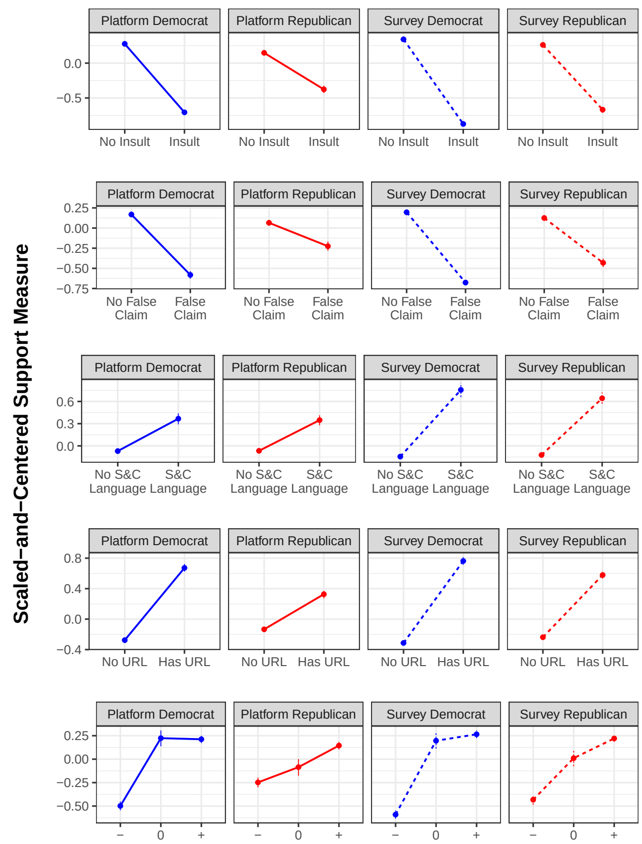

Figure 5: Average value for support measures for tweets according to our independent variables. Each metric is centered and scaled independently to make magnitudes comparable. Thus the y-axis represents change on the scale of a standard deviation of the particular metric. Error bars are 95% bootstrapped CIs. Republican (Democrat) measures are colored red (blue), platform (survey) measures are solid (dotted) lines, and labeled in the grey titles above each subplot. From the top, patterns for whether or not a tweet 1) contains an insult, 2) contains a falsehood, 3) contains support or condolence ("S&C") language, 4) contains a URL, and 5) has negative, neutral, or positive sentiment.

measures. Republicans and Democrats show higher support for non-insulting tweets, tweets that do not contain a falsehood, tweets containing support or condolence language, tweets that have URLs, and tweets with positive sentiment, relative to negative sentiment. The only exception to this rule is Democratic support for neutral tweets, which is no different from their support for positive tweets. Thus, a portion of the agreement across partisan groups is associated with somewhat superficial factors - whether or not a tweet contains a link, for example, or simply whether or not Trump is civil.

Second, while directionality is consistent, the magnitude of effects varies across partisan lines. In particular, measures of Republican support vary less than Democrat support, and

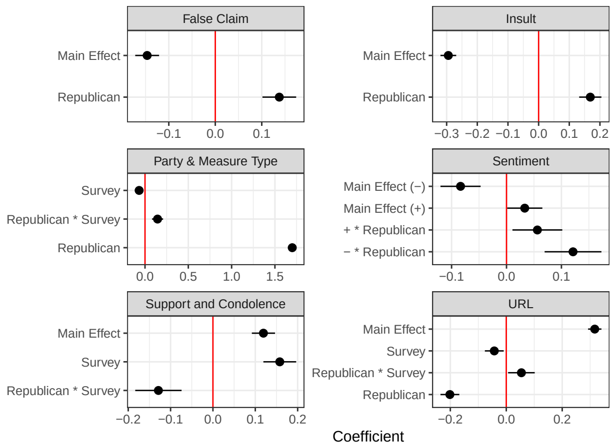

Figure 6: Regression results for our general support model. Coefficients for each variable we consider, along with interaction effects for party and measure type, are shown in separate subplots. There is also a subplot for party and measure factors alone. Shown only are coefficients where standard errors at 95% levels of confidence do not cross 0.

Twitter-based measures vary less than survey-based measures. Note that data in Figure 5 is scaled and centered; Republican metrics are in all cases much higher than Democratic metrics. Consequently, the lack of change across our independent variables suggests that Republicans, especially as measured on Twitter, are more forgiving of Trump's actions, and potentially less willing to show negative support regardless of the content of his tweets. Further, it suggests that individuals may be more likely to react to the content of tweets (at least along the lines considered here) in survey data, relative to behavioral data from social media. Alternatively, it is also possible that self-report data is simply more sensitive to differences across tweets than platform-based behavioral metrics.

We now turn to our regression model to further explore these patterns. The adjusted R-squared of the model is 89.1%, indicating a strong fit to the data. However, it should be noted that some of this explanatory power comes simply from separating Democrats from Republicans. Figure 6 presents results in separate subplots for each of the five explanatory variables, plus main effects for party and measure type. We present all variables where their 95% confidence interval does not cross 0.

Main effects of each of the independent variables aligns with observations from Figure 5. However, in several cases, there are significant interactions with party. Republicans across both measure types rate tweets with false claims, tweets with insults, tweets with negative or positive (relative to neutral) sentiment, and tweets with no URL higher than Democrats. These effects are substantial. For example having a false claim decreases support by .29 standard deviations across all measures [-.32,-.27]. However for Republicans, this effect is diminished by .14 standard deviations [.10,.17], accounting for over half the magnitude of the aggregate effect.

There are fewer, but still noticeable, differences between the two types of metrics. This is partly due to the fact that to leverage a single regression model, we rescaled each measure type by one standard deviation. Consequently, differences in spread noted in Figure 5 across measure type are unobservable in the regression. Still, we find that Democrat survey respondents tend to respond considerably more positively to tweets with support and condolence language and slightly less positively to tweets with URLs. In contrast, considering all interaction effects, there are no differences in effects for Republicans across the two measures.

In sum, we find that Democrats and Republicans have polarized "absolute" support for Trumps tweets. However, this support is correlated on a relative basis - tweets that Republicans think are better (worse) are generally the same as those that Democrats think are better (worse). We show five factors that imply reasons for this agreement, and our findings are largely consistent across two different measures of support. However, we also find that while Democrats and Republicans tend to agree on the general direction of impact of these factors, they differ in the magnitude of impact on their support. Specifically, Republican support, especially as measured by on-platform behavior, is less responsive to content differences in tweets.

Personal Insult Explanatory Model

Figure 7 displays results from our regression model for person insults. Republicans across both measures showed significantly higher support when the target of the insult was a Democrat, relative to a Republican. Compared to an insult of a Republican, an insult of a Democrat was rated .88 standard deviations higher by Republicans on surveys and .62 standard deviations higher on the platform measure. Democrats, in contrast, did not react more positively to Republicans being insulted. In fact, these measures showed a potential for a statistically significant increase in support when Trump insulted either a Democrat (for the platform measure) or someone without an obvious party affiliation (survey measure). However, these observations are not consistent across both measures. Results therefore indicate that Trump insulting members of his own party only serves to hurt himself, relative to insulting a Democrat. Doing so decreases support from his own party, while no obvious, consistent effects occurred when Trump attacked Democrats.

In contrast, the model provides evidence that across both measures, Democrats most strongly disapproved of attacks on members of the media, and on women. These factors had no statistically significant impact on Republicans on either measure of support. Insults of women and media personalities therefore have a net negative impact for Trump, agitating members of the opposing party while not providing a corresponding increase in support from his base. We also find that Democrats in the survey data find insults to Middle Eastern individuals—directed largely at either convicted terrorists or Middle Eastern dictators— less objectionable than insults to white individuals, and find insults of government employees more objectionable than those directed at politicians. In the latter case, however, many of these tweets were directed

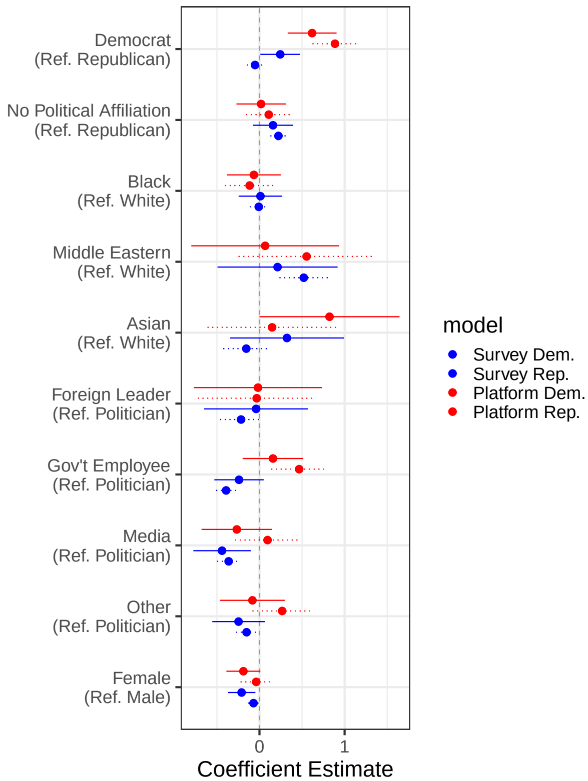

Figure 7: Regression results for the personal insult model for each of the four support measures. Error bars are plus/minus two standard errors. Coefficient estimates are given on the x-axis, model covariates on the y-axis. Coefficients for each of the four different models are given for each variable. Republican (Democrat) measures are colored red (blue), platform (survey) measures are solid (dotted) lines

at James Comey, potentially signifying a more general disapproval of Trump's response to investigations of Russian influence during the 2016 U.S. election.

Overall, both Democrats and Republicans strongly disapproved of Trump's tweets that insult people, there was limited evidence of bipartisan shifts across both support measures depending on the features of the individual insulted. We find only two consistent observations across party lines for both support measures. First, Republican support rose when Trump insulted a Democrat, relative to a Republican, but Democrat support was not similarly (negatively) affected. Second, Democrats responded most negatively when Trump insulted women or members of the media.

Conclusion

The present work investigated partisanship and polarization in response to Donald Trump's tweets. Using both survey- and platform-based measures of support, we found that support for Trump's tweets was highly polarized in absolute terms. However, we also found hidden agreement across party lines. In other words, Republicans and Democrats disagreed on an absolute scale but generally agreed on a relative scale as to which tweets were better and which were worse. This observation holds for both platform- and survey-based measures of support.

In a world where political differences are stressed and partisan attacks are rapidly becoming the norm, this observation of relative agreement is both surprising and encouraging. The fact that it was consistent across two different measures gives us further confidence in its validity. However, this agreement could be partially explained by a simple desire for civility and objectivity - for example, both Democrats and Republicans did not approve of insulting people or telling lies, and agreed that highly positive tweets were good and that highly negative tweets were bad. Consequently, further study is needed to understand the extent to which the observed agreement may lie solely on a superficial level, masking ideological constructs along which we do observe *true* relative polarization.

We also found that in general, Democratic support varied more than Republican support, signaling that Republicans are less responsive to differences across tweets. Republicans also showed more support for insults of Democrats than insults of Republicans, whereas Democratic support is not significantly affected when Trump insults a Democrat relative to a Republican. These observations accord with prior work suggesting that, especially for Republicans, political identity plays an important role in shaping political support (Barber and Pope 2018).

Our work also has broader implications for the measurement of public opinion, suggesting the utility of leveraging complementary measures of public opinion. We found that survey-based and platform-based measures differed in several respects, from factors that explained polarization to estimates of Republican support for negative tweets from Trump. We also found utility in using survey-based measures to tune and validate a platform measure. Using both metrics allowed us greater confidence in certain findings, while also helping to identify potential differences in how support manifests on platforms versus in survey data.

Finally, it is likely important in the future to look beyond the Democrat/Republican divide. This takes two forms. First, it may be useful to move towards more fine-grained representations of political stance. For example, given the split within the Republican party over Trump, examining tweet endorsement from Trump's Republican supporters and non-supporters would be an interesting next step. Second, although data limitations with the aggregate survey measure prevent exploration of it here, it is possible that variables correlated with political party (e.g. age and gender) have equal or greater impacts that political affiliation itself on support for Trump.

In sum, while the concept of polarization generally tends to focus on the existence (or lack thereof) of bimodal distributions of absolute support (Fiorina and Abrams 2008), our work calls for further inquiry into an additional and novel definition of polarization using multiple measures of public support.

References

Ansolabehere, S., and Schaffner, B. F. 2014. Does survey mode still matter? Findings from a 2010 multi-mode comparison. *Political Analysis* 22(3):285–303.

Asur, S., and Huberman, B. A. 2010. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference On*, volume 1, 492–499. IEEE.

Azman, N.; Millard, D. E.; and Weal, M. J. 2012. Dark retweets: Investigating non-conventional retweeting patterns. In *International Conference on Social Informatics*, 489–502. Springer.

Barber, M., and Pope, J. C. 2018. Does party trump ideology? Disentangling party and ideology in America. *American Political Science Review* 1–17.

Barberá, P. 2016. Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data.

Beauchamp, N. 2017. Predicting and interpolating state-level polls using twitter textual data. *American Journal of Political Science* 61(2):490–503.

boyd, D.; Golder, S.; and Lotan, G. 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *Hawaii International Conference on System Sciences* 0:1–10.

Converse, P. E. 1964. The nature of belief systems in mass publics (1964). *Critical review* 18(1-3):1–74.

Davidson, T.; Warmsley, D.; Macy, M.; and Weber, I. 2017. Automated hate speech detection and the problem of offensive language. *arXiv preprint arXiv:1703.04009*.

DellaPosta, D.; Shi, Y.; and Macy, M. 2015. Why Do Liberals Drink Lattes? *American Journal of Sociology* 120(5):1473–1511.

Doherty, C. 2017. Key takeaways on americans growing partisan divide over political values.

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

Fiorina, M. P., and Abrams, S. J. 2008. Political polarization in the american public. *Annu. Rev. Polit. Sci.* 11:563–588.

Friedman, J.; Hastie, T.; and Tibshirani, R. 2009. glmnet: Lasso and elastic-net regularized generalized linear models. *R package version* 1(4).

Garimella, K.; Weber, I.; and De Choudhury, M. 2016. Quote RTs on Twitter: Usage of the New Feature for Political Discourse. In *Proceedings of the 8th ACM Conference on Web Science*, WebSci '16, 200–204. New York, NY, USA: ACM.

Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; and Lazer, D. 2019. Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425):374–378.

Guerra, P. C.; Souza, R. C.; Assunçao, R. M.; and Meira Jr, W. 2017. Antagonism also Flows through Retweets: The Impact of Out-of-Context Quotes in Opinion Polarization Analysis. *arXiv preprint arXiv:1703.03895*.

Guerrero-Solé, F., and Lopez-Gonzalez, H. 2017. Government Formation and Political Discussions in Twitter: An Extended Model for Quantifying Political Distances in Multiparty Democracies. *Social Science Computer Review* 0894439317744163.

Hobbs, W.; Friedland, L.; Joseph, K.; Tsur, O.; Wojcik, S.; and Lazer, D. 2017. "Voters of the Year": 19 Voters Who Were Unintentional Election Poll Sensors on Twitter. In *ICWSM*.

Hutto, C. J., and Gilbert, E. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International AAAI Conference on Weblogs and Social Media*.

Iyengar, S., and Westwood, S. J. 2015. Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science* 59(3):690–707.

Johnson, K., and Goldwasser, D. 2016. Identifying Stance by Analyzing Political Discourse on Twitter. *NLP+ CSS 2016* 66.

Keeter, S.; Hatley, N.; Kennedy, C.; and Lau, A. 2017. What low response rates mean for telephone surveys. Report, Pew Research Center.

Kunda, Z. 1990. The case for motivated reasoning. *Psychological bulletin* 108(3):480.

Lee, J. C., and Quealy, K. 2016. The 487 People, Places and Things Donald Trump Has Insulted on Twitter: A Complete List. *The New York Times*.

Levendusky, M. S., and Malhotra, N. 2016. (Mis)perceptions of Partisan Polarization in the American Public. *Public Opinion Quarterly* 80(S1):378–391.

Levendusky, M. 2009. *The partisan sort: How liberals became Democrats and conservatives became Republicans*. University of Chicago Press.

Lord, C. G.; Ross, L.; and Lepper, M. R. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology* 37(11):2098.

Macskassy, S. A., and Michelson, M. 2011. Why do people retweet? anti-homophily wins the day! In *ICWSM*.

Mason, L. 2015. i disrespectfully agree: The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science* 59(1):128–145.

McGregor, S. C.; Lawrence, R. G.; and Cardona, A. 2017. Personalization, gender, and social media: Gubernatorial candidates' social media strategies. *Information, Communication & Society* 20(2):264–283.

Metaxas, P.; Mustafaraj, E.; Wong, K.; Zeng, L.; O'Keefe, M.; and Finn, S. 2015. What do retweets indicate? results from user survey and meta-review of research. In *Ninth International AAAI Conference on Web and Social Media*. Citeseer.

Monroe, B. L.; Colaresi, M. P.; and Quinn, K. M. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* 16(4):372–403.

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. *Icwsm* 11(122-129):1–2.

Robinson, D. 2016. Text analysis of Trump's tweets confirms he writes only the (angrier) Android half.

Roeder, O.; Mehta, D.; and Wezerek, G. 2017. The Worst Tweeter In Politics Isn't Trump. *FiveThirtyEight*.

Stanyer, J. 2008. ELECTED REPRESENTATIVES, ONLINE SELF-PRESENTATION AND THE PERSONAL VOTE:Party, personality and webstyles in the United States and United Kingdom. *Information, Communication & Society* 11(3):414–432.

Swire, B.; Berinsky, A. J.; Lewandowsky, S.; and Ecker, U. K. 2017. Processing political misinformation: comprehending the trump phenomenon. *Royal Society Open Science* 4(3):160802.

Twyman, J. 2008. Getting it right: YouGov and online survey research in Britain. *Journal of Elections, Public Opinion and Parties* 18(4):343–354.

Vaisey, S. 2014. Is interviewing compatible with the dual-process model of culture. *American Journal of Cultural Sociology* 2(1):150–158.

YouGov. 2016. YouGov TweetIndex · A daily rating of President Trump's tweets. https://tweetindex.yougov.com/.