Brevity is the soul of Twitter: The constraint affordance and political discussion

Kokil Jaidka

Wee Kim Wee School of Communication and Information

Nanyang Technological University

Alvin Y. Zhou

Annenberg School for Communication

University of Pennsylvania

Yphtach Lelkes*

Annenberg School for Communication

University of Pennsylvania

Author Note

Corresponding Author: Yphtach Lelkes, ylelkes@upenn.edu

Abstract

Many hoped that social networks would allow for the open exchange of information and a revival of the public sphere. Unfortunately, conversations on social media are often toxic and not conducive to healthy political discussion. Twitter, the most widely used social network for political discussions, doubled the limit of characters in a Tweet in November 2017, which provided a natural experiment to study the causal effect of technological affordances on political discussions with a discontinuous time series design. Using supervised and unsupervised natural language processing methods, we analyze 358,242 Tweet replies to U.S. politicians from January 2017 to March 2018. We show that the doubling the permissible length of a Tweet led to more polite, less informal, more analytical, and overall healthier discussions online. However, the declining trend in the political relevance of these tweets raises concerns about the implications of the changing norms for the quality of political deliberation.

Keywords: Political Communication, Political Discussion, Social Media, Computational Social Science

Brevity is the soul of Twitter: The constraint affordance and political discussion

Over the past decades, we have witnessed a massive change in the democratic potential of the internet. In the context of political campaigning, social media platforms such as Twitter and Facebook now provide politicians with a platform for outreach to mobilize their voter base and engage in meaningful discussions with their constituents (Theocharis, Barberá, Fazekas, Popa, & Parnet, 2016). At the same time, these platforms give citizens the ability to access political information, participate in local events and citizen movements, and voice their support or dissent against their government. Social networks are potentially important tools in the democratic and civic process, as they offer a neutral and open platform for dialogue, act as the bedrock where citizens build ideas about politics through everyday talk (Wojcieszak & Mutz, 2009; Wyatt, Katz, & Kim, 2000), and facilitate direct contact between citizens and their political representatives (e.g., Liu & Zhang, 2013; Tufekci & Wilson, 2012).

The effectiveness of social networks as platforms for democratic discourse depends on whether or not they are used for polite and respectful discussions (e.g., Liu & Zhang, 2013; Mutz & Reeves, 2005; Papacharissi, 2004). However, political discourse online is often toxic, provocative, and replete with trolls who incite hyperbole (e.g., Berry & Sobieraj, 2013; Nithyanand, Schaffner, & Gill, 2017; Theocharis et al., 2016). Many explanations for the low quality of online discourse have been put forth, including the existence of disinformation campaigns, the polarization of politics, spurious accounts and self-perpetuating echo chambers (e.g., Berry & Sobieraj, 2013; Chen, 2017). Incivility ultimately influences the motivation of other users to have meaningful and constructive exchanges with each other or with their political representatives (Theocharis et al., 2016).

We argue that one factor that regulates the quality of political discussions are the technological affordances for communication. Affordances are properties that enable or constrain the potential for action (Faraj & Azad, 2012). Examples of affordances in digital communication include the ability to have real-time discussions, to hide one's true identity

and to navigate information in a hierarchical manner (Friess & Eilders, 2015; Sundar, 2008). Some technological affordances, we argue, may affect the quality of democratic discussion. For instance, a number of studies have shown that the effect of one affordance (anonymity) on degrading the quality of online political discussions (e.g., Halpern & Gibbs, 2013; Theocharis, Lowe, van Deth, & García-Albacete, 2015; Towne & Herbsleb, 2012).

In this paper, we examine whether another affordance—constraints on message length—impacts the quality of online political discussion. Using an interrupted times series design combined with an automated content analysis of over three hundred and fifty thousand tweets involving politics, we compare various message features associated with ideal political discussion before and after Twitter moved from allowing 140 characters per tweet to 280 characters. We show that Twitter's relaxation of its message length constraint affected some, but not all messages features which improved the quality of political discussion. For instance, while the changeover increased the prevalence of analytical, polite, and formal messages, there were no significant changes in the offensiveness, swearing or cognitive markers. We also test the robustness of our results using different model specifications and bandwidths.

Our theoretical contribution is twofold. We extend prior research by shifting attention to social media affordances that ultimately influence the quality of discussions on social media. Changes to the design of platforms, which are more tractable than, for instance, censoring content, have substantive effects on the health of online political discussion. Second, by revealing the trade-offs in content and style that users face when articulating their opinion, our study yields important insights about how the use of these platforms may affect the quality of public discourse and users' experiences of online political participation as a more reactive versus a more deliberative exercise.

The Twittersphere

Political discussions on Twitter and other online platforms allow people with diverse perspectives and opinions to participate in a common conversation In theory, they can cut

across diverse social networks and deconstruct echo chambers of influence and opinion through the free and open dissemination of messages. In addition to facilitating horizontal discussion between citizens, platforms like Twitter allow vertical discussions between citizens and policymakers. (e.g., Ausserhofer & Maireder, 2013; Davis, 2010; Effing, Van Hillegersberg, & Huibers, 2011).

Twitter's users organically create a networked sphere of political discussion which is structurally independent of the traditional arena of politics or news; yet, it connects with the two through official affiliations and real-life interactions (Lindgren & Lundström, 2011). The online political sphere has the potential to reinvigorate offline politics by allowing millions of individual contributions, subverting the often monolithic agenda set by traditional mass media and policymakers in the offline world (e.g., Habermas, Lennox, & Lennox, 1974; Papacharissi, 2010; Shirky, 2008).

The fact that Twitter facilitates discussion between citizens and policymakers does not, in itself, make it a "democratic utopia" (Papacharissi, 2004; Stroud, Scacco, Muddiman, & Curry, 2015). Online political discussions rarely meet the ideals for political deliberation, such as open communication, equality, inclusivity and compromise (e.g., Friess & Eilders, 2015; Nithyanand et al., 2017; Theocharis et al., 2016; Wojcieszak, 2010). Political discussions on Twitter are often replete with toxic and abusive responses, flaming, and group-based stereotyping (Halpern & Gibbs, 2013; Theocharis et al., 2016).Users participating in discussions on Twitter are found to be unlikely to indulge in reflection, or frame coherent arguments, which negatively impacts the quality of political discussions (Janssen & Kies, 2005; Stromer-Galley & Martinson, 2009). A study by Theocharis et al. (2016) argued that users' toxic and uncivil responses were responsible for shutting down any meaningful political engagement between elected U.S. politicians and the citizenry.

To summarize, many studies have identified a gap in the normative ideals of Twitter as an ideal platform for lively and inclusive political debates, and its role in actually facilitating political deliberation. The following section contextualizes the study's

hypotheses in the current understanding of how Twitter's technological features, or affordances, enable deliberation and politeness or constrain incivility.

Technological Affordances and the Potential for Civic Discussion

When designing online platforms, developers make many choices that change "possibilities for action between an object/technology and the user that enables or constrains potential behavioral outcomes in a particular context" (Evans, Pearce, Vitak, & Treem, 2017, p. 36). These possibilities for action are often called affordances, and can be broadened or narrowed by changing the underlying technical specifications. For instance, in the context of political discussions, Twitter affords the scope for two-way communication between politicians and citizens because of the way in which asymmetrical friendship relationships are specified by default (Grant et al., 2010). The affordance for two-way communication is restricted on Facebook, where, in the default setting, two individual users may need to "follow" each other in order to send a message or even view each other's profile.

Friess and Eilders (2015) and Janssen and Kies (2005) identify the affordances which affect political deliberation. Anonymity affords higher participation but uncivil discourse (Towne & Herbsleb, 2012). Real-time participation in synchronous chats provokes instantaneous reactions but reduces reflexive, coherent argumentation and rebuttal (Janssen & Kies, 2005; Stromer-Galley & Martinson, 2009). Flattened follower-followee connections overcome the depersonalizing effects of digital communication and increase citizens' emotional closeness to political elites and elected candidates (Lee & Oh, 2012).

Navigable information interfaces support political deliberation by fostering clear communication, rational argumentation and constructiveness (Towne & Herbsleb, 2012). Among the studies exploring the role of affordances in political deliberation, a majority have focused on how the anonymity and deindividuation reduces the rationality, sincerity and civility of the conversation (e.g., Halpern & Gibbs, 2013; Theocharis et al., 2015; Towne & Herbsleb, 2012); on the other hand, anonymity also improves the likelihood of

participation and hence the quantity and inclusivity of political debates (Towne & Herbsleb, 2012).

At least two studies have examined the relationship between allowable message length (another affordance, and our particular focus) and political discussion quality. However, these studies turn up conflicting findings: Papacharissi (2004) reported that longer messages posted in political discussions were significantly more uncivil than shorter messages. On the other hand, Oz, Zheng, and Chen (2018) found that shorter messages on Twitter were more uncivil, impolite and less deliberative than longer messages on Facebook; however, in their experimental replication, they observed only a significant improvement in deliberation in the longer messages.

While both studies are important contributions to the literature, they are limited in a number of ways. Papacharissi (2004) does not account for self-selection, and the possibility that those that tend to write longer messages may already behave more uncivilly. While the Oz et al. (2018) experiment solves this internal validity issue, the character limit is confounded with other characteristics and norms of the platform. That is, subjects responded to either a mock Facebook post or a mock Twitter post. Given these conflicting findings, we begin with a research question:

R1. What effect does doubling the character limit have on incivility and politeness?

Character limits may also limit the formality of discourse, making it less amenable for clear and open communication. As the original Twitter users embraced its character limit, they organically devised a series of conventions to convey meaning and information structure: slang, netspeak, and abbreviations were adopted to express reactions (LOL, SMH), share opinion (IMHO, AFAIK), reference other users (-mentions, h/t, via), label topics (hashtags), and identify propagated messages (RT, QT). Accordingly, before the character limit change on Twitter, we expect that users would attempt to squeeze as much information as possible into 140 characters, leading to higher numbers of instances of informality, netspeak, and word complexity.

However, while the need for concision required adherence to convention, this may lead to the sacrifice of style over content. Twitter users are constrained to expressing themselves in 140 characters, which have been found to comparatively reduce self-disclosure and emotions in self-expression as compared to longer posts, and posts on other social media platforms (Lin & Qiu, 2013). Some scholars argue that emotions play a positive role in political discussion: Gastil (2008) and Papacharissi (2004) espouse emotionally charged rhetoric that permits "frank public testimony and moral debate"(Gastil, 2008, p. 65). Accordingly, after the character limit change, we expect that users would be more likely to use the extra characters to add subjectivity to their responses with an affective underscoring (of either positive and negative emotion), hence,

H1: Comments posted on Twitter before the 280-character limit change will be (a) more informal and (b) less emotional than after the change.

A different perspective posed by Chen (2017), and Lovejoy, Waters, and Saxton (2012) and others is that the type-box limitation of Twitter prevents meaningful argumentation. One of the potential benefits of extending the character limit may be that it affords users to more space to make a cogent argument or support their views with evidence, thus facilitating more sophisticated and deliberative political discussions. The study by Oz et al. (2018) reported less deliberation in political messages addressed to the U.S. White House on Twitter as compared to Facebook; in an experimental replication the participants were once again significantly more deliberative in their Facebook vs. their Twitter posts. An improvement in political deliberation need not necessarily be orthogonal to an expectation of greater incivility. Chen (2017) speculates that both incivility and political deliberation can increase when message length constraints are relaxed; however, this proposition has not controlled for anonymity effects. Accordingly, we posit our second hypothesis:

H2: Comments posted on Twitter before the 280-character limit change will be less deliberative than after the change.

Civility is an essential but not sufficient condition for achieving the "democratic potential" of the internet. Political deliberation also needs to broach substantive topics. Civility may be negatively correlated with substantiveness: Papacharissi (2004) raises the concern that greater adherence to politeness could restrict conversation by making it "reserved, tepid, less spontaneous." Instead of simply incivility, political discussion should encompass "a wider range of topics, and conversation specifically aimed at political action" (see also, Freelon, 2010; Halpern & Gibbs, 2013; Theocharis et al., 2016). As we observe a research gap in the link between social media affordances and the substantiveness of messages, we pose the following research question:

R2. What effect does doubling the character limit have on the political relevance of messages?

Data and Methods

We built our dataset of political discussions on Twitter by adopting the method of Theocharis et al. (2016) to identify the Twitter replies to 536 U.S. Congressmen and Congresswomen who were in office before November 7, 2017, when Twitter instituted a change in the character limit. We filtered a Twitter 1% sample, collected using the Twitter streaming API between January 2017 - March 2018, to retain all replies (i.e., tweets starting with "politicians") to these set of Twitter handles, occurring between January 1, 2017 and March 31, 2018. These dates give us a larger pre-intervention and post-intervention observation period and the ability to assess whether our results remain stable if we consider different bandwidths in the regression equation.

In Table 1, we provide a summary of our dataset with the number of tweets remaining after each pre-processing step. First, we performed language filtering to retain only English-language tweets. Next, we removed any tweets which were retweets and thus did not constitute an actual reply. We manually inspected the character limits of the tweets and found that because of the differences in encoding, the actual length of tweets in the pre-intervention period was often up to 145 characters. Accordingly, we considered 145

characters as a better approximation to compliance than 140 characters for future experiments. We removed tweets by the 1% users of the population who were subjected to the intervention early (the 'early-access' users) on September 27 . Finally, our dataset comprises 358,242 replies respectively to U.S. politicians from January 2017 - March 2018.

Table 1

Dataset Description. Standard errors are in parantheses.

Number of	**Replies in the 1% Sample**	**After language filtering**	**After removing retweets**	**After removing early-access**
observations	2101856	1869481	398278	358242
Number of	**Unique users**	**Non-compliers**		**Compliers**
users	204902	62180		39246
Tweet	**Total number of words**	**Average words per tweet**		**Mean daily tweets**
characteristics	7668123	21.40 (12.6)	12.07	772.56 (479.76)

Table 1 also provides statistics on the unique number of users in our dataset. We also identify the *compliers*: users who tweeted more than 145 characters at least once after the character limit change, and the *noncompliers*: users who, despite the extension, never tweeted more than 145 characters. Wilcoxon signed-rank test shows that compliers and noncompliers do not differ significantly as far as their account characteristics, such as the number of followers, the number of followings, the tenure of the account, and whether accounts are verified. However, compliers have a higher average number of tweets posted (μ = 34196.84, SD = 55160.43) and a higher per-user occurrence in our dataset (μ = 2.67, SD=4.80) as compared to non-compliers (μ = 31428.62, SD = 52015.01; μ = 2.03, SD=3.30), which suggests that compliers might be more active on Twiter and already more engaged in politics than noncompliers. A comprehensive comparison of compliers and non-compliers is provided in the Supplementary Materials.

Feature analysis. Based on previous work in analyzing political discourse, civility and deliberation on social media, we have identified a subset of stylistic, emotional,

cognitive and content features in the language that can reflect political deliberation.

Operationalization. All the measures lie on a 0 to 1 scale. The outcomes of the supervised methods used to measure offensiveness and politeness are a binary and a continuous score respectively, and they were calculated using a weighted function of the normalized frequency distribution of linguistic features in a tweet. Outcomes of the unsupervised methods (used to measure the proportion of swear words, informality, affect, deliberation and political relevance) comprise normalized, within-tweet percentages indicating the proportion of words which were representative of the category of interest. For example, a score of 3 for *swear words* would imply that if a tweet comprised a hundred words, then there would be three among those hundred that were swear words.

- **Incivility:** Incivility comprises the use of name-calling, profanity, hate speech or invocation of stereotypes of a homophobic, racist, sexist or xenophobic nature (Chen, 2017). We have applied and compared three different operationalizations of incivility:
  - First, we trained an *offensiveness* classifier developed by Davidson, Warmsley, Macy, and Weber (2017), on a dataset of hand-annotated tweets which are labeled as offensive (1) or inoffensive (0). An example of an uncivil message in our dataset was "<user> DO YOU WANT TO BE REMEMBERED AS A GIANT Pxxxx. U GERRYMANDERED DISTRICTS 4 LYFE. U DUMB MOTHER Fxxxxx NEED TO SLOW THE ROLL." Offensive tweets make up 2.64% of the dataset, consistent with findings from previous literature (Theocharis et al., 2016).
  - Second, we used the *uncivil words* dictionary hand-annotated from a corpus of New York Times comments by Muddiman, McGregor, and Stroud (2018). An example of a message with a high proportion of uncivil words is "<user> Impeach the Whitehouse Hitler. CRIMINAL CRIMINAL ...". 14.3% of the tweets in our dataset contained uncivil words as compared to the 10% reported by the authors in their dataset of comments.

  - From LIWC, we identify the *swear* category which measures the relative proportion of swear words in text, and found that the mean proportion of swear words per tweet was about 55%.

- **Politeness:** Politeness refers to an adherence to etiquette and an extension of courtesy to your fellow discussants Papacharissi (2004). We used the Stanford *Politeness* Application Programming Interface (API) to a language model provided by Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, and Potts (2013) which scores each tweet on a scale of 0 to 1 for its politeness. An example of an impolite message in our dataset was "<user> Why don't you Republicans stop praying and do something! You're just prostitutes for the NRA."; it scored 0.08 in politeness. An example of a very polite message in our dataset was "<user> Please consider my attached letter. I sent it to your colleague, Sen. <name> as well. I appreciate your service...thank you!"; it scored 0.92 in politeness. 24% of tweets in our dataset scored greater than 0.5 on a continuous scale from 0 to 1.

- **Informality**: Informality is defined as content that is relatively higher in its use of assents, fillers, swear words and netspeak. As such, informal messages likely carry less substance, and less effective means of political discussion. We measure informality using LIWC's *informality* variable which summarizes informality along the five dimensions mentioned above, and its subcategory *netspeak* which measures the use of abbreviations, slang and emojis that are characteristic of informal communication on Twitter.

  We also use LIWC's operationalization of *six letter words* which is a count of the number of words with six letters or more. Text with longer words is considered more linguistically dense (Flesch, 1948), which is ideal to abstract opinions into ideas, and compress more information into a 140-character post. On the other hand, text expressing the same meaning with shorter words is regarded to be closer to the actual

meaning, easier to understand, coherent, and thus useful in a political discussion than the former (Cohn, Mehl, & Pennebaker, 2004; J. Pennebaker, Booth, Boyd, & Francis, 2015). Direct political messages with simple language and smaller words have been found to be more effective, persuasive and memorable than ones with more complex messaging (Cobb & Kuklinski, 1997).

- **Affect**: Affect refers to the degree of positive or negative emotion reflected in the text. Higher positive emotion may reflect more amicable political discourse; higher negative emotion with higher anger or sadness could indicate uncivil discourse. We have measured emotional tone using the following dictionaries from LIWC:
  - *Positive emotion*, e.g., "<user> Beautiful honorarium! Thanks!" scored 0.75 on a scale from 0 to 1.
  - *negative emotion*, e.g., "<user> Pathetically Weak" scored 0.67.
  - *Anger*, e.g., "<user> Ban assault weapons!!!!!!!" scored 0.5.
  - *Sadness*, e.g., "<user> Incredibly disappointed." scored 0.33.
- **Deliberation**: Deliberation is considered one of the most important aspects of a democratic society (e.g., Gutman & Thompson, 1996) and is observed in conversation through the use of logic, evidence and rational arguments to make claims and promote the exploration of solutions through dialogue (e.g., Gastil, 2008). In political discourse, deliberation has been found to affect political knowledge, efficacy, and willingness to participate in politics (Min, 2007). In order to measure deliberation, we identify LIWC's *analytical thinking* and *cognitive processing* categories as relevant to our study. A higher score in analytical thinking indicates a formal style of thinking, with more logical and hierarchical arguments; a lower score reflects more informal, personal, here-and-now, and narrative thinking (J. Pennebaker et al., 2015). A tweet with high analytical thinking is "<user> Anti-immigrant is un-American.

Oppose racism/sexism/hate. OPPOSE Sessions for AG! Demand better for America! #StopSessions."

Narrative thinking is also captured through the cognitive processing category, where a higher score reflects the author's tendency to 'think out loud' in their writing through the use of causal or critical language (e.g., think, know, because, should, but, and maybe). A tweet with high cognitive processing is "<user> Can't make any promises but I'll try."

- **Political relevance:** We operationalize the political relevance of a tweet as a measurement of the proportion of its words which are related to politics. We used the political terms dictionary created by Preoţiuc-Pietro, Liu, Hopkins, and Ungar (2017) to score each tweet according to the proportion of *politics words* (e.g., governor, fbi, nra, and congress) and its subcategories, *political entities* (e.g., reid, barack, trump, and biden) and *media entities* (e.g., nbcnews, colbert, wsj, and huffpost).

Validation. We use supervised machine learning models trained on hand-annotated data developed by computer scientists (Danescu-Niculescu-Mizil et al., 2013; Davidson et al., 2017). The classifier and the annotated dataset on offensiveness has been validated in subsequent studies (Almeida, Souza, Nakamura, & Nakamura, 2017; Olteanu, Talamadupula, & Varshney, 2017). The politeness API has also been extensively used by the computer science research community (e.g., Althoff, Danescu-Niculescu-Mizil, & Jurafsky, 2014; Jongeling, Sarkar, Datta, & Serebrenik, 2017; Tan, Niculae, Danescu-Niculescu-Mizil, & Lee, 2016) in linguistic analyses of social media text.

We also use unsupervised methods comprising the psycholinguistic dictionaries provided by Linguistic Inquiry and Word Count (LIWC) 2015 (J. W. Pennebaker, Boyd, Jordan, & Blackburn, 2015), a computerized linguistic program developed by psychologists to automatically categorize words in a text (J. W. Pennebaker et al., 2015), and dictionaries of uncivil words and political words provided by Muddiman et al. (2018) and

Preoţiuc-Pietro et al. (2017). Dictionaries of LIWC have been validated in subsequent language analyses of social media posts; Tausczik and Pennebaker (2010) for more details about validation, and J. W. Pennebaker et al. (2015) for details on LIWC's construction and inter-coder reliability. Further details about the supervised and unsupervised methods used to mine these features are provided in the Supplementary Materials.

Interrupted Time Series (ITS) Regression Model

To determine whether the character-limit change induced an improvement in political discussions, our primary approach is an interrupted time series analysis, a variation of regression discontinuity designs (RDD) in which the running variable is time (Bernal, Cummins, & Gasparrini, 2017). This approach is ideal for our analysis because Twitter data is time-stamped, with a high frequency of daily measurements and a well-defined moment of intervention. ITS design requires a clear differentiation of the pre-intervention and post-intervention period (Bernal et al., 2017)—the extension of character limit from 140 to 280 on November 7, 2017 in our case. The unit of analysis in this set of regressions is the daily mean score of a linguistic dimension. All variables were mean-centered before analysis. To tackle the inconsistency of the numbers of tweets on each day and avoid Type I errors, we bootstrapped the analysis for 100 iterations, sampling an approximate daily mean of 700 tweets per day for each analysis and report the average effect sizes and standard errors across all the iterations. The quantity of interest is the immediate change in overall incivility, politeness, informality, affect, deliberation and political content after November 7, 2017, represented by t in the following ordinary least squares models:

$$Y_{feature,t} = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t \quad (1)$$

In the above model, T is the relative time distance from November 7, 2017. For example, *time* equals to -3 on November 4, 2017 and equals to 7 on November 14, 2017. X is a dummy variable indicating whether the tweets were published before (coded 0) or after the intervention (coded 1). Therefore, $beta_2$ indicates the intercept shift following the

character limit intervention on the feature value, while b_3 shows the slope change after the character limit intervention. Including quadratic or cubic time trends did not substantively change our results (see Supplementary Materials). Using this regression model, we examined the effect of character limit intervention on various linguistic outcomes $Y_{feature}$. The regression results in tables show the average of the 100 iterations, while the LOESS figures show one random iteration for purposes of illustration. We focus on the 100 days before and 100 days after the character switch (i.e., our bandwidth and N=200).

In addition to following the standard ITS design to estimate the impact of the intervention, we also present a simple difference in means in the feature prominence pre and post- the intervention for expository purposes. We also compare the pre-intervention levels among the entire sample to post-intervention levels only among tweets that contained more than 145 characters (compliers) and among those that contained less than 145 characters (non-compliers).[1] We expect the effect of the intervention to have only changed the language characteristics of compliers.

Our study is similar to a regression discontinuity design, wherein the intervention day can be thought of as a discontinuity and time as the running variable (Hausman & Rapson, 2017). Since not every person "complies" with treatment (uses more than 145 characters), we use a fuzzy regression discontinuity framework, which treats the number of characters in a tweet as an endogenous variables (coded 0 if less than 145 characters; 1 if more than 145 characters), and a dummy indicating whether a person tweeted before (coded 0) or after (coded 1) November 7 as the instrumental variable. That is, the intervention exogenously increases the probability someone uses more than 145 characters. Our causal estimand in this model is the Local Average Treatment Effect, i.e., the effect among those who used more than 145 characters.

Finally, we tested our results for evidence of a self-selection bias, wherein people who are more likely to tweet different types of messages are more likely to use more than 145

[1] We model this using the same framework as the ITS model.

characters. To assuage these concerns, we limited the dataset to those subjects that tweeted before and after the intervention and included subject fixed effects, which effectively examines the effect of the character limit change within subjects, washing away time-invariant effects.

Results

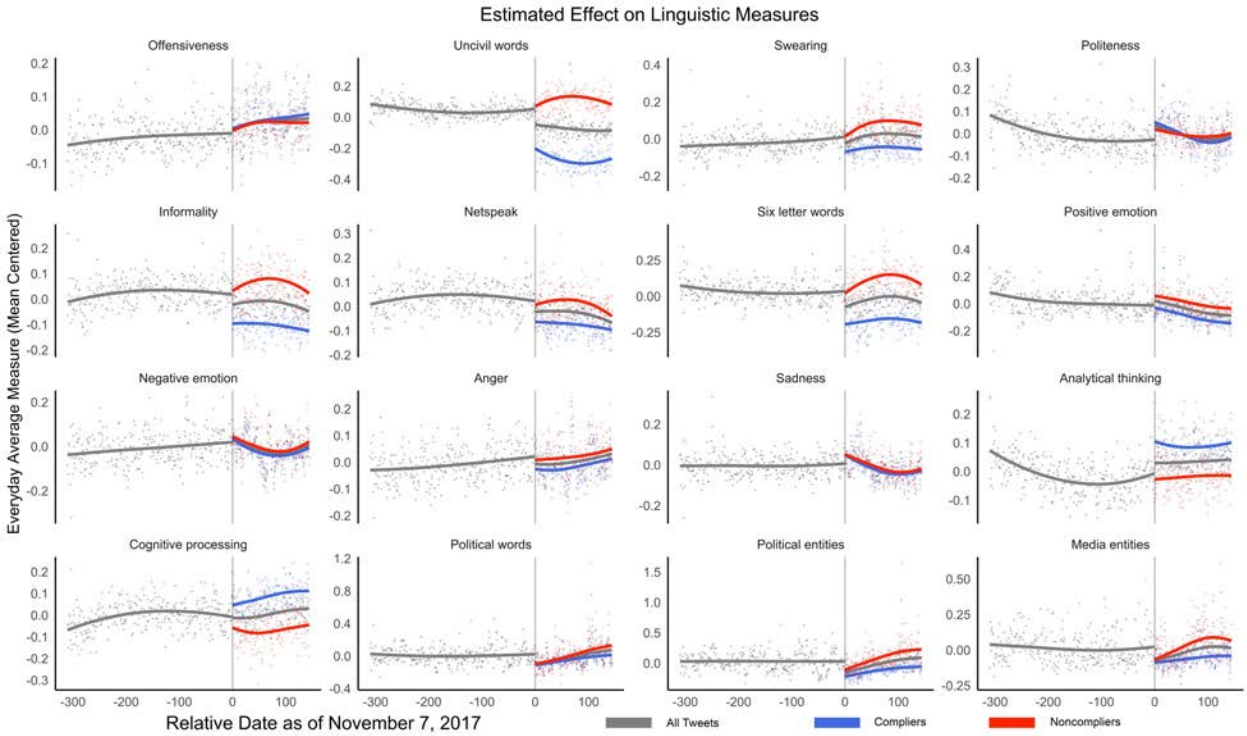
As a manipulation check, we first assessed whether intervention affected the number of words used per tweet. The average number of words per tweet increasing significantly after the character limit intervention (β: .47, p<.001). This effect did not decline over time (coefficient of interaction: .001, p>.05).

The core analysis is performed on 358,242 tweets, of which 146,878 were posted after the intervention (Twitter's character limit change). Compliers posted 70,040 tweets (47.7%) after the intervention. Figure 1 shows the change in linguistic dimension in the period leading up to and after the intervention, where the solid curves were generated by locally weighted regression of the linguistic dimensions on sequential day numbers, with no adjustment for covariates. Post-intervention, the blue curve reflects the trend for compliers while the red curve depicts the non-compliers.

This visualization provides the first set of clues regarding the effect of the intervention on online political discourse. Politeness, positive emotion and sadness appear to have increased discontinuously on the day of the intervention but adhere to the overall downward trend. Swear words, informality and netspeak decreased discontinuously on the day of the intervention but the slope remains relatively level; anger, too, decreased discontinuously but follows the overall rising trend thereafter. No discontinuities at the intervention are observed for offensiveness and overall negative emotion. Post-intervention, compliers (blue) and non-compliers (red) appear to be markedly different in their use of swear words, informality, netspeak, use of six letter words, positive emotions and anger.

We formally test these difference in Table 2. These OLS estimates are fit to the daily

Figure 1. Daily trends of linguistic dimensions from January 2017 to March 2018 for political discussions on Twitter. The solid curve is generated by locally weighted (LOESS) regression of the linguistic dimensions on sequential day numbers, with no adjustment for covariates. Data was aggregated as day-level means here to facilitate visualization.



trends for a bandwidth of 100 days and feature two functional forms.[2] Aggregating by day, Column 1 presents an estimate of the mean difference in the linguistic dimensions before and after the intervention. Column 2 fits a linear function to the data on either side of the intervention date. Columns 3 and 4 provide the results considering only compliers and non-compliers respectively, post-intervention. Column 5 provides the local average treatment effects among compliers following a fuzzy regression discontinuity design implementation. Columns 6 and 7 report the estimates after considering subject fixed effects, for those in our dataset who tweeted both pre- and post-intervention.

[2] While separate models were used to fit quadratic and cubic functions to the data, those models did not substantially increase model fit. Those results have been provided in the Supplementary Materials.

Table 2

OLS Estimates of the effect of platform change with different model and treatment specifications (bandwidth = 100 days)

	1 Difference in Means	2 Linear	3 Linear Compliers	4 Linear Non-compliers	5 IV Estimates	6 Within All Subjects	7 Within Subject: Compliers
Incivility							
Offensiveness	.039***	.026	.014	.018	.08	.03·	.02
	(.009)	(.018)	(.019)	(.02)	(.05)	(.018)	(.021)
Uncivil words	-.04***	-.03***	-.09***	.01*	-.15***	-.04***	-.08***
	(.003)	(.006)	(.006)	(.007)	(.018)	(.001)	(.007)
Swearing	.02 ·	-.02	-.08***	.03	-.12*	-.00+	-.01
	(.009)	(.019)	(.018)	(.023)	(.05)	(.017)	(.021)
Politeness	.04***	.07***	.08***	.05*	.31***	.06***	.04*
	(.009)	(.019)	(.022)	(.019)	(.05)	(.017)	(.021)
Informality							
Informality	-.04***	-.03	-.12***	.03	-.31***	-.02	-.02
	(.009)	(.019)	(.019)	(.021)	(.05)	(.017)	(.021)
Netspeak	-.06***	-.05*	-.09***	-.01	-.36***	-.04**	-.04*
	(.009)	(.019)	(.019)	(.021)	(.06)	(.017)	(.021)
Six letter words	-.05***	-.09***	-.22***	.02	-.74***	-.08***	-.07***
	(.012)	(.025)	(.027)	(.026)	(.05)	(.162)	(.204)
Affect							
Positive emotion	-.02	.02	-.01	.08**	.19***	.04***	.03·
	(.013)	(.025)	(.025)	(.028)	(.05)	(.016)	(.020)
Negative emotion	-.02·	.02	.00+	.02	.15**	.05***	.05***
	(.010)	(.023)	(.023)	(.025)	(.017)	(.021)	(.067)
Anger	-.01	-.00+	-.05*	-.00+	-.09·	-.00+	-.02
	(.009)	(.018)	(.021)	(.023)	(.06)	(.017)	(.021)
Sadness	-.00+	.02	.02	.03	.12*	.04**	.04·
	(.009)	(.018)	(.018)	(.019)	(.06)	(.017)	(.020)
Deliberation							
Analytical thinking	.07***	.04*	.09***	-.02	.06	.00+	.00+
	(.009)	(.018)	(.020)	(.02)	(.06)	(.017)	(.021)
Cognitive processing	-.02*	-.03	.02	-.09***	-.07	-.02	-.016
	(.009)	(.019)	(.018)	(.023)	(.05)	(.017)	(.021)
Political content							
Political words	-.003***	-.004***	-.004***	-.004***	-.021***	-.005***	-.004***
	(.0005)	(.0018)	(.001)	(.001)	(.003)	(.001)	(.001)
Political entities	-.004***	-.005***	-.005***	-.005***	-.027***	-.005***	-.006***
	(.0005)	(.001)	(.001)	(.001)	(.002)	(.001)	(.001)
Media entities	-.001***	-.002***	-.002***	-.002***	-.01***	-.002***	-.002***
	(.0002)	(.0003)	(.0003)	(.0003)	(.0009)	(.0003)	(.0003)
Overall mentions	.03***	.03	.15***	-.06***	.06		
	(.01)	(.02)	(.02)	(.02)	(.05)		
Observations	200	200	200	200	180634	15271	7382

Note: ·p<.1;*p<.05; ** p<.01; *** p<.001
Standard errors are shown in parentheses.

To answer our first research question, we observe that three out of four measures of incivility changed post-intervention to make the discussion more civil overall. Offensiveness only differed in the first (and least credible) specification. The use of uncivil words significantly decreased among compliers (Column 3: β = -.09, p<.001; Column 5: β = -.15, p<.001; Column 7: β = -.08, p<.001), but it showed a significant increase among non-compliers (Column 4: β = .01, p<.05). While the overall use of swear words did not significantly change, it significantly decreased among the compliers (Column 3: β = -.08, p<.001; Column 5: β = -.12, p<.05; Column 7: β = -.01, p>.1). Politeness significantly increased (β = .07, p<.001) across all model specifications. While politeness also increased among non-compliers, the effect was twice as large among the compliers than non-compliers (Column 3: β = .08, p<.001; β = .05, p<.05).

Language informality decreased after the character-limit change. All three dimensions showed a significant decrease among the compliers as can be seen in columns 3 (-.22 < β < -.09, p<.001) and 5 (-.74 < β < -.31, p<.001). To summarize, there was a significant decrease in the use of uncivil words, netspeak, and six letter words by compliers, and a significant increase in politeness, and these effects were significant at the within-subject level. Therefore, *H1a is supported by the results.*

There were no consistent results across the model specifications among the results for affect.The local average treatment effect for the compliers were strong, significant and positive for positive emotion, negative emotion and sadness (Column 5: .12 < β < .19, p<.001) and they were corroborated by the weaker and marginally significant within-subject fixed effects (Column 7: .03 < β < .05, p <.1). Overall, *H1b is partially supported by the results.*

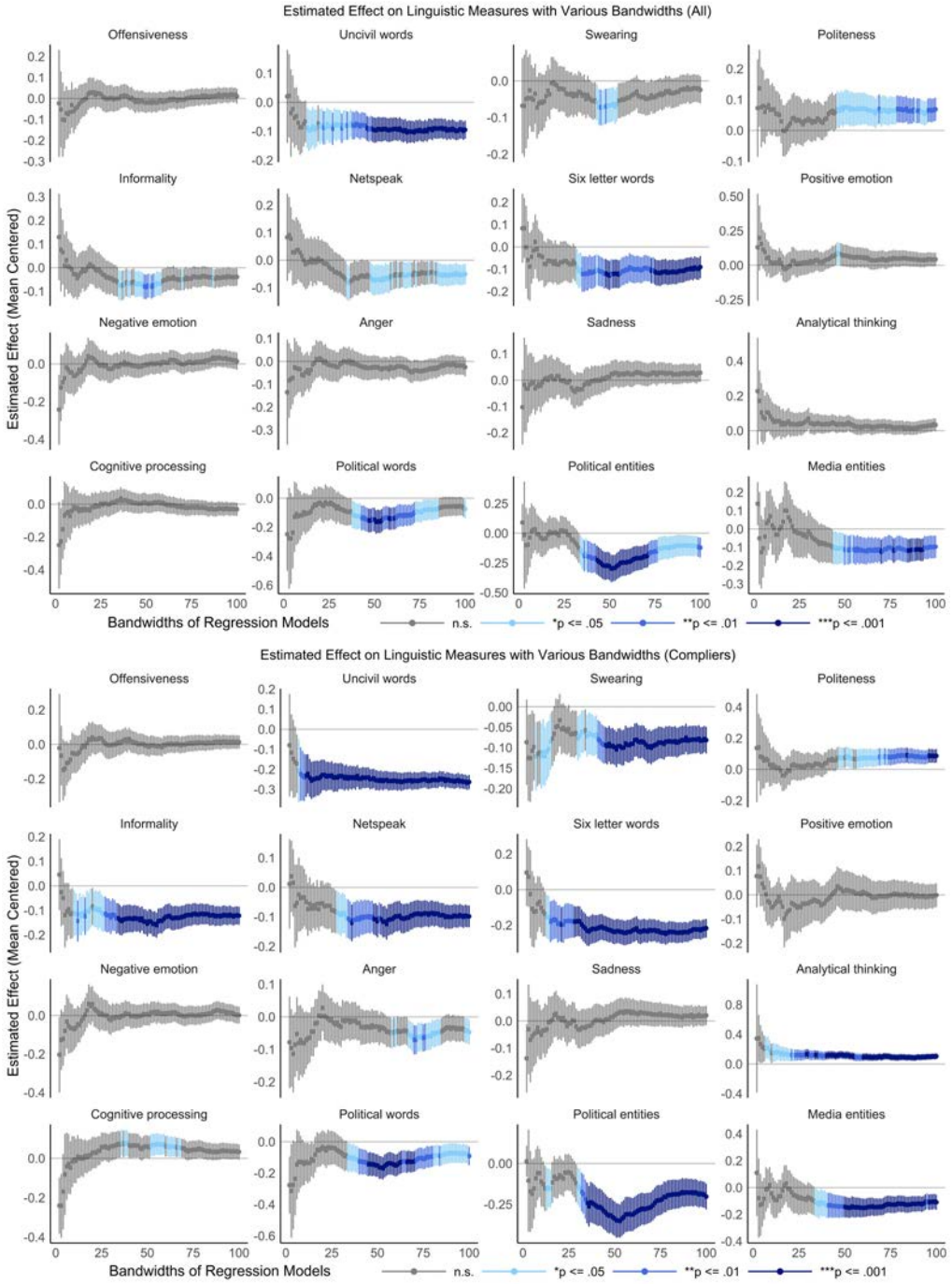
Finally, the results for deliberation – analytical thinking and cognitive processing – lie in opposite directions, which is not unexpected, because they measure contrasting thinking styles. The overall increase in analytical thinking is significant among the

compliers (Column 1: β = .07, p<.001; Column 2: β = .04, p<.001 Column 3: β = .09, p<.001), indicating that their writing became more logical, hierarchical and reasoning-oriented. On the other hand, cognitive processing shows a significant decrease among the non-compliers (Column 4: β = -.09, p<.001), indicating that their writing became less tentative, casual and narrative-oriented. Analytical thinking does not show a significant local average treatment effect or within-subject fixed effect, but the coefficient is similarly sized and in the same direction as in the other models, giving us confidence in the estimates from Columns 1, 2, and 3. Therefore *H2 is supported by the results.*

In order to answer R2, we consider the changes in the political relevance of the tweets. From Table 2, we see that both compliers and non-compliers showed a significant drop in the mentions of political words (Columns 3 and 4: -.004< β <-.002, p<.001) and its subcategories political entities (-.03< β <-.005, p<.001) and media entities (-.01< β <-.002, p<.001). These coefficients are the strongest as the local average treatment effect among the compliers (Column 5: -.027< β <-.01, p<.001) and are also replicated among the within-subject fixed effects (Column 7: -.006< β <-.002, p<.001). To summarize, our findings suggest that the compliers were less likely to make political references, and the overall conversation was likely less focused on politics than before the character limit change. We further contextualize the apparent drop in political content by also examining the effect of the intervention on the overall number of -mentions. The results are provided in the last row of Table 2. We found that (a) the drop in the mentions of political and media entities by compliers was opposite to a general, significant positive effect in the number of mentions (β = .15, p<.001); (b) non-compliers, on the whole, were significantly less likely to make *any* mentions after the platform change (β = -.06, p< .001).

To interpret the effect of the character limit change on the political discussion, we note that since the treatment only takes the value 0 or 1, the standardized coefficients indicate the linear increment or decrement to the within-tweet proportion of each dependent variable 'y' post-intervention. This would imply that the compliance effect

Figure 2. Estimates of the average change in linguistic dimensions after the intervention by fitting a linear model over (a) all observations (N = 358,242) and (b) compliers (N = 281,804), using various bandwidths [0,100] as plotted on the x-axis. Vertical lines denote 95% confidence intervals for each estimate.



(Column 3) for Twitter users ranged from a per-tweet decrement of .2% (for mentions of media entities) and as much as 22% (for six letter words). We note then that the post-intervention differences for political content are small at the tweet level and the subject level (Columns 5 and 7 of Table 2). Nevertheless, we expect that across millions of tweets and people, these small effects lead to a large difference in the overall characteristics of the political conversation, as can be visualized in the apparent discontinuity in Figure 1.

Robustness and Placebo Tests

To test whether our results were an artifact of the bandwidth chosen in our ITS models, we replicated our models with different bandwidths of data. Figure 2(a) displays the estimated treatment effects using the linear functional form with all observations for bandwidth N = [1,100], and Figure 2(b) provides them when considering only compliers post-treatment. Shaded points indicate that the estimates are significant at different levels ($.01 < p < .001$). These trends show that while the choice of bandwidth affects the significance of effects when there is a smaller sample size, the direction of the effect remains consistent across large and small bandwidths; hence, our findings do not appear to be contingent on the choice of bandwidth.

We also tested whether our results were an artifact of the measurement method, because our dependent variables are a function of the number of words in a tweet. Because all feature measurements were normalized percentages calculated at the tweet level; thus, they are objectively independent of length. Nevertheless, as an additional validation check we reran our analysis on a subset of 54,217 tweets that were about the same length (i.e. 135 - 155 characters). Our findings are available in the Supplementary Materials, and confirm the trends reported in Table 2.

We replicated our analysis on a dataset of tweets posted in 2016, using the same data collection method as above to first identify 66,927 replies to the same U.S. politicians in the 100-day bandwidth before and after November 7, 2016. We then pre-processed our dataset to obtain the linguistic features and bootstrapped our experimental analysis for a

difference in means and a linear model specification. We report the average standardized coefficients and standard errors for 100 iterations, conducted on an average of 400 daily observations sampled with replacement from the dataset. The results are reported in the Supplementary Materials and, as hoped, they do not repeat the main trends of Table 2 post- the placebo treatment.

Discussion and Recommendations

While many past studies have examined the relationship of various affordances to political discussions, these studies have been limited to cross-sectional observational settings or artificial lab settings. The current study takes advantage of the Twitter character limit change and employs a quasi-natural experiment approach, thus strengthening previous scholars' arguments specifically about online political deliberation, and more generally about the importance of affordances in the study of computer-mediated communication.

There are certainly many operationalizations of incivility, and we offer a comprehensive examination of three different measures to cover the different facets of uncivil behavior. Communication scholars have argued that incivility is more than offensiveness. For example, Papacharissi (2004) provided an early definition of incivility as impolite behavior that threatens democracy and has lasting repercussions on the common good. In other words, uncivil tweets must be offensive, but offensive tweets are not necessarily uncivil. Papacharissi's (2004) conceptualization was adopted by recent studies using human coding (e.g., Groshek & Cutino, 2016; Muddiman et al., 2018; Rowe, 2015; Santana, 2014). Across the different incivility measures, we show that our findings are consistent, and distributionally similar to those reported in other corpora of political comments by Theocharis et al. (2016), Oz et al. (2018) and Muddiman et al. (2018).

As with all quasi-experimental designs, we cannot discount all threats to internal validity (Cook, Campbell, & Shadish, 2002). In particular, what Cook et al. (2002) call internal validity threat due to history remains possible. November 8, 2017 is the one-year

anniversary of the 2016 United States presidential election, which is very close to the date of Twitter character limit change. The event might have briefly changed the linguistic characteristics of the tweets around that time. However, the effect would eventually fade away, while the new 280 character limit would not. And as our analysis shows, the effect of the character limit change on certain linguistic measures remains months after the event. Furthermore, a placebo test conducted using the data of November 2016 reaffirms the validity of our conclusions. We have also tested our results for evidence of a self-selection bias and found that our main results were replicated in this condition; however, not all the analyses had statistical power when only 7000 compliers were being considered. Nevertheless, the coefficient is similarly sized and in the same direction as in the other models.

The findings of our study have practical implications for future work. Firstly, our analysis led us to discover an overall significant positive effect in the number of -mentions in tweets following the character limit change. -mentions can be used to include a greater diversity of voices into the conversation, thus potentially changing the speed of information dissemination, and the organic structure of the networks that emerge in Twitter discussions. Twitter's new algorithms allow users to see tweets mentioning their followers; thus, by the simple act of adding a few extra characters, we speculate that users can potentially get their message out to thousands of second- and third-degree connections. Future work could explore whether the extra length affordances could help to break through the tightly polarized communities existing in the general Twittersphere.

Secondly, although we employed the Twitter streaming API in our study, which only returns a 1% random sample of all public tweets, we recommend scholars to use Twitter's firehose stream when possible, which charges based on the date range and the number of tweets requested by researchers. Studies have debated on the sampling validity of the streaming API for quite some time (see González-Bailón, Wang, Rivero, Borge-Holthoefer, & Moreno, 2014; Morstatter, Pfeffer, & Liu, 2014; Morstatter, Pfeffer, Liu, & Carley, 2013,

for details). However, a time-series analysis requires a wide band of data, which often makes it monetarily unrealistic to use the full-access, yet expensive, alternative.

Finally, we recommend that the limitations of the sampling issue and the historic threat could be tested by extending this research project to other contexts. We invite colleagues to replicate our study in other countries and languages to examine whether the effect of affordance change on political deliberation is contingent on cultural and political factors. Further research can also be conducted in experimental settings to mitigate environmental confounders and further our understanding of the dynamics between technological affordances and human communication.

Conclusion

The increasing accessibility of larger datasets and the development of new tools for computational and linguistic analyses have enabled scholars to use more sophisticated techniques to model computer-mediated communication; this also has larger implications for the study of political communication. Methodological frameworks to identify causal effects, such as instrumental variables and regression discontinuity designs, are also becoming increasingly popular in social science disciplines (e.g., Dunning, 2008). We hope that this study encourages others in communication to combine computational approaches with the tools of causal inference to identify media effects in the digital world.

Many have argued that popular social media platforms might not be suitable for sophisticated and serious discussions, citing the negative human behavior demonstrated online such as toxicity, incivility, and lack of empathy (e.g., Anderson, Brossard, Scheufele, Xenos, & Ladwig, 2014; Baek, Wojcieszak, & Delli Carpini, 2012; Coe, Kenski, & Rains, 2014). Some have suggested that a solution to online incivility lies in exposing people to different content or exhortations to act civil (e.g., Kim, 2015; Munger, 2017). This study suggests that examining technological affordances are also a way forward. We present a quasi-experiment on the impact of a technological affordance change on political deliberation. Twitter's implementation of a character limit change to double the length of

tweets led to more polite, less informal, and more analytical political discussions online.

While the character limit improved overall civility in political deliberation, we do have two major concerns about its implications. Firstly, political discussions appear to be less substantive post the change, with fewer political references. The effect is robust across various regression models and bandwidths. Our findings lead us back to the debate among political deliberation scholars: incivility in political deliberation may not necessarily be a negative quality. Political messages with incivility are often considered more entertaining (Mutz & Reeves, 2005), more memorable (McGuire, 1981), more persuasive (Fridkin & Kenney, 2008), and more authentic (Benson, 2011) than a polite, well-formed argument, perhaps signaling sincerity and the speaker's high stakes in the disagreement (Benson, 2011). The findings also support the concerns raised by previous scholars about the importance of 'disagreement and anarchy' in the online political sphere for true democratic emancipation (e.g., Papacharissi, 2004; Schudson, 1997).

A second concern is that politeness as an emerging norm may be responsible for creating structural inequalities in the online public sphere. To begin with, our findings show that the effects of Twitter's new affordance is largely limited to those who self-select into applying them in their daily communication. Therefore, the overall improvement in the civility and analytical nature of political discussions on Twitter may only matter for those social media users who are already politically engaged. Scholars have shown that elite users—already more engaged, articulate and authoritative—lead the political rhetoric in the *dominant*, rather than the public sphere of debate (e.g., Fraser, 1990; McGregor, 2018; Papacharissi, 2004).

We conclude with the implications of our study for social media's role in political deliberation. Through this study, we have shown how the design decisions made by computer engineers and technology companies might have a profound impact on democratic processes. In light of recent events where communication technology had a detrimental effect on democracy, such as the cybersecurity issues around the 2016 US

presidential election, rampant misinformation on social media (Del Vicario et al., 2016), and some evidence of online echo chambers (at least on social media) (e.g., Barberá, Jost, Nagler, Tucker, & Bonneau, 2015; Flaxman, Goel, & Rao, 2016; Sunstein, 2017), we contend that technology companies have a responsibility to make communication platforms more friendly for political discussions. However, this too, should be implemented with careful consideration of users' rights; this study does not advocate technological determinism as a catch-all solution for improving the online political sphere.

To summarize, while an online political utopia may remain a pipe-dream, this study has shown that we can make the current environment more hospitable to democratic discourse. While platforms are understandably cautious when it comes to censoring or promoting certain content or users, this study indicates that they have other tools that will promote healthy political discourse. Internet platforms are constantly A/B testing their products to increase profits and revenue, and to improve user engagement and satisfaction (Hindman, 2018). They may also want to test the impact of such design changes on discussion health. Of course, platforms will need to balance their profit motives for their public interest motives when making these decisions, but we hope that the latter will play a significant role in this calculus.

References

Almeida, T. G., Souza, B. À., Nakamura, F. G., & Nakamura, E. F. (2017). Detecting hate, offensive, and regular speech in short comments. In *Proceedings of the 23rd brazillian symposium on multimedia and the web* (pp. 225–228).

Althoff, T., Danescu-Niculescu-Mizil, C., & Jurafsky, D. (2014). How to ask for a favor: A case study on the success of altruistic requests. In *Proceedings of the international aaai conference on web and social media.*

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387.

Ausserhofer, J., & Maireder, A. (2013). National politics on twitter: Structures and topics of a networked public sphere. *Information, Communication & Society*, *16*(3), 291–314.

Baek, Y. M., Wojcieszak, M., & Delli Carpini, M. X. (2012). Online versus face-to-face deliberation: Who? Why? What? With what effects? *New Media & Society*, *14*(3), 363–383.

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

Benson, T. W. (2011). The rhetoric of civility: Power, authenticity, and democracy. *Journal of Contemporary Rhetoric*, *1*(1).

Bernal, J. L., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology*, *46*(1), 348–355.

Berry, J. M., & Sobieraj, S. (2013). *The outrage industry: Political opinion media and the new incivility*. Oxford University Press.

Chen, G. (2017). *Online incivility and public debate: Nasty talk*. Springer.

Cobb, M. D., & Kuklinski, J. H. (1997). Changing minds: Political arguments and political persuasion. *American Journal of Political Science*, 88–121.

Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, *64*(4), 658–679. doi: 10.1111/jcom.12104

Cohn, M. A., Mehl, M. R., & Pennebaker, J. W. (2004). Linguistic markers of psychological change surrounding september 11, 2001. *Psychological science*, *15*(10), 687–693.

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Boston.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A

computational approach to politeness with application to social factors. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)* (Vol. 1, pp. 250–259).

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media.*

Davis, A. (2010). New media and fat democracy: the paradox of online participation1. *New media & society*, *12*(5), 745–761.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559.

Dunning, T. (2008). Improving causal inference: Strengths and limitations of natural experiments. *Political Research Quarterly*, *61*(2), 282–293.

Effing, R., Van Hillegersberg, J., & Huibers, T. (2011). Social media and political participation: are facebook, twitter and youtube democratizing our political systems? In *International conference on electronic participation* (pp. 25–35).

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A aonceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52.

Faraj, S., & Azad, B. (2012). The materiality of technology: An affordance perspective. *Materiality and organizing: Social interaction in a technological world*, *237*, 258.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320.

Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, *32*(3), 221.

Fraser, B. (1990). Perspectives on politeness. *Journal of pragmatics*, *14*(2), 219–236.

Freelon, D. G. (2010). Analyzing online political discussion using three models of democratic communication. *New media & society*, *12*(7), 1172–1190.

Fridkin, K. L., & Kenney, P. J. (2008). The dimensions of negative messages. *American Politics Research*, *36*(5), 694–723.

Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, *7*(3), 319–339.

Gastil, J. (2008). Mediated deliberation and public opinion. in: Gastil, john. *Political Communication and Deliberation. Sage Publications*.

González-Bailón, S., Wang, N., Rivero, A., Borge-Holthoefer, J., & Moreno, Y. (2014). Assessing the bias in samples of large online networks. *Social Networks*, *38*, 16–27. doi: 10.1016/j.socnet.2014.01.004

Groshek, J., & Cutino, C. (2016). Meaner on mobile: Incivility and impoliteness in communicating contentious politics on sociotechnical networks. *Social Media + Society*, *2*(4), 205630511667713. doi: 10.1177/2056305116677137

Gutman, A., & Thompson, D. (1996). Democracy and disagreement: Why moral conflict cannot be avoided in politics, and what should be done about it. *Cambridge, MA: Belknap*.

Habermas, J., Lennox, S., & Lennox, F. (1974). The public sphere: An encyclopedia article (1964). *New German Critique*(3), 49–55.

Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, *29*(3), 1159–1168. doi: 10.1016/j.chb.2012.10.008

Hausman, C., & Rapson, D. S. (2017). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*(0).

Hindman, M. (2018). *The internet trap: How the digital economy builds monopolies and undermines democracy*. Princeton University Press.

Janssen, D., & Kies, R. (2005). Online forums and deliberative democracy. *Acta política*, *40*(3), 317–335.

Jongeling, R., Sarkar, P., Datta, S., & Serebrenik, A. (2017). On negative results when

using sentiment analysis tools for software engineering research. *Empirical Software Engineering*, *22*(5), 2543–2584.

Kim, Y. (2015). Does disagreement mitigate polarization? how selective exposure and disagreement affect political polarization. *Journalism & Mass Communication Quarterly*, *92*(4), 915–937.

Lee, E.-J., & Oh, S. Y. (2012). To personalize or depersonalize? When and how politicians' personalized tweets affect the public's reactions. *Journal of Communication*, *62*(6), 932–949.

Lin, H., & Qiu, L. (2013). Two sites, two voices: Linguistic differences between facebook status updates and tweets. In *International conference on cross-cultural design* (pp. 432–440).

Lindgren, S., & Lundström, R. (2011). Pirate culture and hacktivist mobilization: The cultural and social protocols of #WikiLeaks on Twitter. *New Media & Society*, *13*(6), 999–1018.

Liu, N., & Zhang, X. (2013). The influence of group communication, government–citizen interaction, and perceived importance of new media on online political discussion. *Policy & Internet*, *5*(4), 444–461.

Lovejoy, K., Waters, R. D., & Saxton, G. D. (2012). Engaging stakeholders through Twitter: How nonprofit organizations are getting more out of 140 characters or less. *Public Relations Review*, *38*(2), 313–318.

McGregor, S. C. (2018). *Social (media) construction of public opinion by elites* (Unpublished doctoral dissertation).

McGuire, W. J. (1981). Theoretical foundations of campaigns.

Min, S.-J. (2007). Online vs. face-to-face deliberation: Effects on civic engagement. *Journal of Computer-Mediated Communication*, *12*(4), 1369–1387.

Morstatter, F., Pfeffer, J., & Liu, H. (2014). When is it biased?: Assessing the representativeness of Twitter's Streaming API. In *Proceedings of the 23rd*

International Conference on World Wide Web - WWW '14 Companion (pp. 555–556).

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter's Streaming API with Twitter's Firehose. In *Proceedings of the International AAAI Conference on Web and Social Media* (pp. 400–408).

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2018). (re)claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *0*(0), 1-13.

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649.

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(01).

Nithyanand, R., Schaffner, B., & Gill, P. (2017). Online political discourse in the trump era. *arXiv preprint arXiv:1711.05303*.

Olteanu, A., Talamadupula, K., & Varshney, K. R. (2017). The limits of abstract evaluation metrics: The case of hate speech detection. In *Proceedings of the 2017 acm on web science conference* (pp. 405–406).

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, *20*(9), 3400–3419.

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283.

Papacharissi, Z. (2010). *A private sphere: Democracy in a digital age.* Malden, MA: Polity.

Pennebaker, J., Booth, R., Boyd, R., & Francis, M. (2015). *Linguistic inquiry and word count: Liwc 2015 [computer software]. pennebaker conglomerates.* Inc.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development*

and psychometric properties of LIWC2015 (Tech. Rep.). Austin, TX: University of Texas at Austin.

Preoţiuc-Pietro, D., Liu, Y., Hopkins, D., & Ungar, L. (2017). Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (Vol. 1, pp. 729–740).

Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, *18*(2), 121–138.

Santana, A. D. (2014, January). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, *8*(1), 18–33.

Schudson, M. (1997). Why conversation is not the soul of democracy. *Critical Studies in Mass Communication*, *14*(4), 297–309.

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations*. New York, NY: Penguin Press.

Stromer-Galley, J., & Martinson, A. M. (2009). Coherence in political computer-mediated communication: Analyzing topic relevance and drift in chat. *Discourse & Communication*, *3*(2), 195–216.

Stroud, N. J., Scacco, J. M., Muddiman, A., & Curry, A. L. (2015). Changing deliberative norms on news organizations' Facebook sites. *Journal of Computer-Mediated Communication*, *20*(2), 188–203.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). Cambridge, MA: MIT Press.

Sunstein, C. R. (2017). *#Republic: Divided democracy in the age of social media*. Princeton, NJ: Princeton University Press.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In

Proceedings of the 25th international conference on world wide web (pp. 613–624).

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A bad workman blames his tweets: The consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of Communication*, *66*(6), 1007–1031.

Theocharis, Y., Lowe, W., van Deth, J. W., & García-Albacete, G. (2015). Using Twitter to mobilize protest action: Online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements. *Information, Communication & Society*, *18*(2), 202–220.

Towne, W. B., & Herbsleb, J. D. (2012). Design considerations for online deliberation systems. *Journal of Information Technology & Politics*, *9*(1), 97–115.

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from tahrir square. *Journal of communication*, *62*(2), 363–379.

Wojcieszak, M. (2010). 'Don't talk to me': Effects of ideologically homogeneous online groups and politically dissimilar offline ties on extremism. *New Media & Society*, *12*(4), 637–655.

Wojcieszak, M., & Mutz, D. (2009, March). Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement? *Journal of Communication*, *59*(1), 40–56. Retrieved 2018-04-25, from `https://academic.oup.com/joc/article/59/1/40-56/4098524`

Wyatt, R. O., Katz, E., & Kim, J. (2000). Bridging the spheres: Political and personal conversation in public and private spaces. *Journal of communication*, *50*(1), 71–92.

Supplementary Materials: Brevity is the soul of Twitter: The constraint affordance and political discussion

Kokil Jaidka
Wee Kim Wee School of Communication and Information
Nanyang Technological University

Alvin Y. Zhou
Annenberg School for Communication
University of Pennsylvania

Yphtach Lelkes*
Annenberg School for Communication
University of Pennsylvania

Author Note

Corresponding Author: Yphtach Lelkes, ylelkes@upenn.edu

Contents

Covariate Balance check: pre-intervention differences between compliers and non-compliers

We conducted Mann-Whitney U tests on the characteristics of compliers and non-compliers prior to the intervention to ascertain whether there were differences prior to the intervention. Compliers were more politically engaged than non-compliers to begin

with. Non-compliers were more likely to use swear words and express negative emotions than compliers, before the intervention. More details can be found in Table S1.

Differences in meta-characteristics

Prior to the intervention, we observed that there were more tweets per complier than tweets per non-complier in our 1% sample, suggesting that compliers were already more politically engaged than non-compliers. Compliers also tweeted more than non-compliers and had more words per sentence and greater deliberative characteristics (higher analytical thinking, higher cognitive processing and higher use of political words) than non-compliers to begin with. Non-compliers were more likely to use uncivil words and to swear, as compared to compliers.

Differences in linguistic characteristics

- In terms of incivility, non-compliers were significantly more likely to use swear words as compared to compliers, suggested by the offensiveness category and the uncivil words category.
- Non-compliers were more likely to express negative emotion before the intervention than compliers.
- In terms of deliberation, compliers are higher than non-compliers pre intervention.

The following sections describe each of the language models applied in this study in further detail.

Supervised Machine Learning Models

Offensiveness classifier. We used the code provided by Davidson, Warmsley, Macy, and Weber (2017) [1] to train an offensiveness classifier on the original

[1] GitHub repository at https://goo.gl/vGXM5H

Table S1

Prior to the intervention, compliers and non-compliers were significantly different in terms of their engagement with politicians, number of tweets, and some linguistic measures. Values are mean. p values are based on the Mann-Whitney U test.

	Compliers	Non-compliers	p
Whether verified	0.22%	0.25%	.64
Number of tweets	34196.84	31428.62	***
Number of following	2162.97	1986.40	.64
Number of followers	2268.20	2229.30	·
Tenure (years, as of 11/2017)	4.18	4.21	.40
Number of tweets per user in dataset	1.67	1.19	
Words per sentence	12.73	11.42	***
Offensiveness	1.91%	2.21%	·
Uncivil words	0.13%	2.60%	***
Swearing	0.44	0.58	·
Politeness	0.45	0.45	·
Informality	3.60	3.52	***
Netspeak	2.73	2.42	***
Six letter words	29.62	29.82	.40
Positive emotion	3.09	3.27	.16
Negative emotion	2.90	2.93	**
Anger	1.20	1.30	.17
Sadness	0.49	0.47	**
Analytical thinking	60.89	58.17	***
Cognitive processing	8.18	8.07	**
Political words	4.92%	4.90%	*
Political entities	2.30%	2.41%	.33
Media entities	0.43%	0.45%	*
Note:	·p<.1;*p<.05; ** p<.01; *** p<.001		

hand-annotated dataset provided as a part of the paper. We then applied it to our dataset to generate labels for offensiveness at the tweet level.

Feature extraction. The original classifier used a number of features which are functions of tweet length, such as the Flesch Kincaid reading ease scores; however, we discarded any such features to ensure that our classifier was not inadvertently tautological. First, the classifier was trained on the labeled training set. Unlike the original paper, we decided to use a balanced dataset by undersampling the negative class. This was done to ensure higher classification accuracy per tweet. Feature extraction was done as follows:

- All the text was lowercased and stemmed using the Porter stemmer to create features representing words as unigrams, pairs of words as bigrams and triplets of words as trigrams.

- The presence of words was converted into a frequency distribution, and weighted according to their importance in the tweet by calculating their term frequency - inverse document frequency (TF*IDF).

- Syntactic features were created using the Stanford Natural Language Toolkit (NLTK) (Manning et al., 2014) to label the unigrams, bigrams and trigrams according to their Penn Part-of-Speech (POS) tags.

- Sentiment features (positive, negative, neutral and compound sentiment) were calculated using the VADER package in NLTK.

We trained the classifier using regularized logistic regression with L2 penalty as provided by the python library *scikit-learn*, and the final classifier obtained an accuracy of 96.7% in five-fold cross-validation on held out data.

Next, we transformed our tweet dataset into the same set of linguistic features as mentioned above. We then applied the trained classifier, which generated a 0/1 label for each tweet signifying whether it was not/was offensive.

Table S2

Model coefficients for the offensiveness classifier

Weighted TF-IDF features	
Feature	Coefficient
'a**'	-.32
'bird'	.05
'b**ch'	-.72
'browni'	.04
'charli'	.04
'c**t'	-.27
'da'	-.05
'f**'	-.25
'fa***t'	-.31
'f**k'	-.43
'h**'	-.49
'nic**'	-.17
'ni*'	-.19
'nig**'	-.38
'ni****'	-.29
'ni****'	-.28
'ni****'	-.25
'pu***'	-.44
'retard'	-.18
'sh*t'	-.29
'trash'	.03
'white trash'	-.18
'yanke'	.06
'yellow'	.04
Part of Speech N-grams	
'NN IN'	.02
'NN NN NN'	.05
Sentiment features	
VADER neutral sentiment	.21
VADER compound sentiment	.06
Meta-features	
Number of hashtags	.003
Number of at-mentions	.01

The Stanford Politeness API. The Politeness API was trained using supervised machine learning on 10,000 annotated utterances culled from Wikipedia and StackExchange (Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, & Potts, 2013). The model extracts linguistic features which embody politeness, such as indirection, deference, impersonalization and modality, and demonstrated an accuracy of up to 83% on the original corpus. We used the code provided by the original authors [2] to preprocess and scored the tweets in our dataset on their politeness.

Unsupervised Methods

LIWC dictionaries. Informality, affect and deliberation were measured using the LIWC 2015 dictionaries, which comprise over 6400 words and word stems. The English version of LIWC has demonstrated high internal reliability and external validity for its different emotional and cognitive measurements (for more details, see Pennebaker, Boyd, Jordan, & Blackburn, 2015). LIWC has extensively been used in communication literature (e.g., Correa, Scherman, & Arriagada, 2016; Kapidzic & Herring, 2015; Valenzuela, Piña, & Ramírez, 2017), and in the analysis of online discussions of political and social issues (Ahmed, Jaidka, & Cho, 2017; De Choudhury, Jhaver, Sugar, & Weber, 2016) to measure individual differences and aggregate trends in social relationships, thinking styles, and emotional expression (see also Bae & Lee, 2012; Golder & Macy, 2011).

LIWC categorizes each word according to the dictionaries in which its word stem is listed – for example, *cri*, the word stem of *cry*, would be classified into four categories: *affective process*, *negative emotion*, *sadness*, and *verb*. The outcome of the LIWC classifier indicates what percentage of words were categorized into each measure. For example, a score of .03 for *negative emotion* means that for every 100 words analyzed, there are 3 words that contained negative emotion. LIWC generates over 90 output variables, but we focus on the measures of informality ($\alpha = .84$), netspeak ($\alpha = .82$), six letter words and

[2] GitHub repository at https://goo.gl/jFd4Zh

swearing ($\alpha = .83$); affect based on positive emotion ($\alpha = .64$), negative emotion ($\alpha = .55$), anger ($\alpha = .53$) and sadness ($\alpha = .70$), and deliberation based on analytical thinking ($\alpha = .71$) and cognitive processing ($\alpha = .92$), which are related to our concerns about political discussion. [3]

Uncivil words dictionary. The uncivil words dictionary was published by Muddiman, McGregor, and Stroud (2018) in a content analysis of incivility specific to political deliberation, on social media. Following standard approaches, the dictionary was created by annotating the most frequent words in a corpus of 9.6 million comments posted to articles in the New York Times between October 2007 and August 2013, then validating the relevance of individual words as uncivil in comments by using trained annotators. The final dictionary comprises 88 uncivil words with a Krippendorff's α of .87.

Political content dictionary. The political content dictionary was developed by Preoţiuc-Pietro, Liu, Hopkins, and Ungar (2017) in order to compare the group-level differences in the online political engagement of Twitter users. Similar to the uncivil words dictionary, this dictionary was created by following an annotation-based approach followed by a deductive approach to refine a subset of the most frequent 12,000 unigrams, which were extracted from a dataset of 4.8 million Twitter posts contributed by 3,938 survey participants (adults in the United States). We found that 59% of the tweets in our dataset contained political words.

ITS Model Specifications

Effect sizes were estimated using ordinary least squares regression, where the model was specified as either a linear, quadratic or cubic function. This section presents the details of the model specifications and the R^2 values thus obtained. Because sophisticated models provided only a marginal contribution for model explainability, the manuscript

[3] The dimensions we do not examine are grammatical variables and parts of speech, perceptual and biological processes, time orientation, social processes, personal concerns and drives.

consistently uses the linear specification for all the main results.

Linear:

$$Y_{feature,t} = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t \tag{1}$$

Quadratic:

$$Y_{feature,t} = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t + \beta_4 T X_t^2 + \beta_5 X_t^2 \tag{2}$$

Cubic:

$$Y_{feature,t} = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t + \beta_4 T X_t^2 + \beta_5 X_t^2 + \beta_6 T X_t^3 + \beta_7 X_t^3 \tag{3}$$

In the above models, T is the relative time distance from November 7, 2017. For example, *time* equals to -3 on November 4, 2017 and equals to 7 on November 14, 2017. X is a dummy variable indicating whether the tweets were published before (coded 0) or after the intervention (coded 1). Therefore, β_2 indicates the intercept shift following the character limit change on the feature value, while β_3 shows the slope change after the character limit intervention. The quadratic and cubic models show additional variables for the different orders of X and their interaction with T.

Model fit for various functional specifications

In order to justify our choice of a linear functional specification instead of a quadratic or cubic function, we report the ten-fold cross-validation errors in model performance for predicting each of the dependent variables on one-tenth held out data after training it on the remaining nine-tenths of the data. Table S3 provides the mean absolute errors and the mean square errors corresponding to the linear, quadratic, and cubic functions respectively, for data within a bandwidth <= 100 days from the intervention date. As the table shows, across a variety of specifications, quadratic or cubic specifications do not substantially improve the explanatory power of the model.

Table S3

Polynomial order	WPS	Offensive	Uncivil words	Swearing	Politeness	Informality	Netspeak	Six letter words	Positive emotion	Negative emotion	Anger	Sad	Analytical thinking	Cognitive Processing	Media entities	Political entities	Political terms
MAEs																	
0	.67	.37	.0122	.49	.73	.69	.62	.76	.68	.72	.62	.46	.85	.77	.007	.0272	.041
	(.002)	(.005)	(.000$^+$)	(.003)	(.002)	(.002)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)	(.000$^+$)	(.000$^+$)	(.000$^+$)
1	.67	.37	.0122	.4907	.728	.69	.62	.76	.6811	.72	.62	.4671	.85	.77	.0075	.02718	.041
	(.002)	(.005)	(.000$^+$)	(.003)	(.002)	(.003)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)	(.000$^+$)	(.000$^+$)	(.000$^+$)
2	.674	.3746	.0122	.4904	.729	.69	.6232	.7601	.681	.7237	.622	.467	.85	.7714	.0074	.02718	.0411
	(.002)	(.005)	(.000$^+$)	(.004)	(.002)	(.002)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)	(.000$^+$)	(.000$^+$)	(.000$^+$)
3	.6748	.37	.012	.4904	.729	.6905	.623	.7596	.681	.7232	.622	.4668	.85	.7713	.0074	.0271	.041
	(.002)	(.005)	(.000$^+$)	(.004)	(.002)	(.002)	(.003)	(.002)	(.001)	(.002)	(.002)	(.001)	(.002)	(.002)	(.000$^+$)	(.000$^+$)	(.001)
MSEs																	
0	1.54	1.17	.0005	1.09	1.08	.98	.97	1.07	.89	.97	.99	.97	.95	.94	.0003	.0016	.0029
	(.195)	(.035)	(.000$^+$)	(.033)	(.006)	(.013)	(.016)	(.009)	(.009)	(.013)	(.016)	(.015)	(.003)	(.006)	(.000$^+$)	(.000$^+$)	(.000$^+$)
1	1.54	1.169	.000$^+$	1.097	1.08	.98	.97	1.07	.890	.97	.99	.977	.95	.939	.000$^+$	.002	.003
	(.195)	(.035)	(.000$^+$)	(.329)	(.005)	(.013)	(.017)	(.009)	(.009)	(.013)	(.16)	(.015)	(.003)	(.006)	(.000$^+$)	(.000$^+$)	(.000$^+$)
2	1.548	1.169	.000$^+$	1.097	1.080	.989	.968	1.077	.890	.973	.998	.977	.95	9.939	.000$^+$	.002	.003
	(.195)	(.035)	(.000$^+$)	(.003)	(.006)	(.013)	(.017)	(.009)	(.009)	(.013)	(.16)	(.015)	(.003)	(.006)	(.000$^+$)	(.000$^+$)	(.000$^+$)
3	1.5482	1.169	.0005	1.097	1.08	.989	.968	1.075	.890	.972	.997	.977	.953	.939	.000$^+$	.002	.003
	(.195)	(.035)	(.000$^+$)	(.033)	(.006)	(.012)	(.017)	(.009)	(.009)	(.013)	(.016)	(.014)	(.003)	(.006)	(.000$^+$)	(.000$^+$)	(.000$^+$)

Length sensitivity analysis

Table S4 provides the main results on a subset of the data comprising only 54000 tweets which range from 135 to 155 characters in length. We see that the main findings are replicated in this dataset in all cases except analytical thinking.

Placebo test: 7 November 2016

Table S5 provides the results for the difference in means and the linear model specification on a replication of the experiment, conducted as a placebo test by collecting, pre-processing and analyzing 66,927 replies to the same US politicians in a 100-day period before and after 7 November 2016, collected from Twitter's 1% sample. Our expectation was that the lack of an evident intervention by Twitter would lead to non-significant findings from this dataset. The table reports that there were no significant differences reported in politeness, informality or analytical thinking, which are the key findings reported in Table 2. Our offensiveness classifier did not label any tweets as offensive in this dataset. There was, however, a steep drop in the use of political content words after November 7, an effect which is a hundredfold greater than the one reported in Table 2. Therefore, we remain confident in the validity of our results and inferences.

Table S4

Length sensitivity analysis: OLS Estimates of the effect of platform change on linguistic characteristics for tweets that are 135 - 155 characters in length (bandwidth = 100 days)

	Difference in Means	Linear	Compliers	Non Compliers	IV Estimates
			Incivility		
Offensiveness	.03*	-.009	.006	-.024	-.075
	(.016)	(.032)	(.04)	(.042)	(.186)
Uncivil words	-.03***	-.042·	-.059*	-.022	-.33**
	(.010)	(.021)	(.026)	(.028)	(.122)
Swearing	.03 ***	.001	.035	-.039	-.11
	(.011)	(.023)	(.029)	(.028)	(.119)
Politeness	.027	.09*	.11**	.07	.363·
	(.02)	(.039)	(.046)	(.050)	(.203)
			Informality		
Informality	-.109***	-.133***	-.122***	-.140***	-.834***
	(.012)	(.026)	(.035)	(.030)	(.162)
Netspeak	-.15***	-.169***	-.169***	-.167***	-1.020***
	(.013)	(.027)	(.037)	(.034)	(.179)
Six letter words	-.09***	-.077*	-.059	-.095**	-.554***
	(.032)	(.027)	(.043)	(.037)	(.152)
			Affect		
Positive emotion	.032*	.050·	.038	.047	.166
	(.012)	(.026)	(.033)	(.030)	(.129)
Negative emotion	-.04 *	.019	.016	.029	.088
	(.016)	(.032)	(.036)	(.044)	(.151)
Anger	-.038*	-.023	.001	-.047	-.199
	(.014)	(.028)	(.032)	(.039)	(.141)
Sadness	-.01	.039	.032	.050	.213
	(.016)	(.031)	(.035)	(.037)	(.155)
			Deliberation		
Analytical thinking	-.039*	-.012	-.046	.006	-.05
	(.017)	(.034)	(.040)	(.041)	(.169)
Cognitive processing	.037*	-.032	-.046	-.027	-.256·
	(.015)	(.029)	(.038)	(.039)	(.155)
			Political content		
Political words	-.05***	-.059*	-.068*	.07	-.469**
	(.015)	(.030)	(.046)	(.051)	(.149)
Political entities	-.07***	-.08*	-.09**	-.083*	-.86***
	(.0017)	(.032)	(.037)	(.037)	(.141)
Media entities	-.027*	-.01	.004	-.040	-.211·
	(.012)	(.024)	(.029)	(.031)	(.113)
Observations	200	200	200	200	24851

Note: ·p<.1;*p<.05; ** p<.01; *** p<.001
Standard errors are shown in parentheses.

Table S5

Placebo test: OLS Estimates for a replication analysis conducted on tweets collected from 2016, noting the effect of a non-treatment on November 7, 2016, on a change in the linguistic characteristics (bandwidth = 100 days)

	Difference in Means	Linear	Within Subjects
	Incivility		
Offensiveness	–	–	–
Uncivil words	.03·	.02	.04
	(.016)	(.036)	(.057)
Swearing	.05***	.09*	.09·
	(.016)	(.094)	(.052)
Politeness	.07***	.07	.06
	(.020)	(.050)	(.052)
	Informality		
Informality	-.08***	-.059	-.02
	(.017)	(.045)	(.052)
Netspeak	-.14***	-.09	-.08
	(.020)	(.050)	(.051)
Six letter words	-.06*	.01	-.038
	(.029)	(.063)	(.054)
	Affect		
Positive emotion	.103***	.006	.12
	(.019)	(.043)	(.024)
Negative emotion	.05*	-.02	-.03
	(.019)	(.032)	(.054)
Anger	.05*	.06	.06
	(.017)	(.038)	(.056)
Sadness	.01	-.02	-.06
	(.018)	(.040)	(.055)
	Deliberation		
Analytical thinking	-.09***	.10·	.01
	(.027)	(.057)	(.052)
Cognitive processing	.01***	.02	.11*
	(.020)	(.045)	(.054)
	Political content		
Political words	-.22***	-.17***	-.13*
	(.029)	(.063)	(.054)
Political entities	-.28***	-.19*	-.15***
	(.035)	(.075)	(.052)
Media entities	.007	-.01	-.04
	(.019)	(.043)	(.058)
Observations	200	200	1056

Note: ·p<.1;*p<.05; ** p<.01; *** p<.001
Standard errors are shown in parentheses.

References

Ahmed, S., Jaidka, K., & Cho, J. (2017). Tweeting india's nirbhaya protest: a study of emotional dynamics in an online social movement. *Social Movement Studies*, *16*(4), 447–465.

Bae, Y., & Lee, H. (2012). Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society for Information Science and Technology*, *63*(12), 2521–2535.

Correa, T., Scherman, A., & Arriagada, A. (2016). Audiences and disasters: Analyses of media diaries before and after an earthquake and a massive fire. *Journal of Communication*, *66*(4), 519–541.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. In *Proceedings of the 51st annual meeting of the association for computational linguistics (volume 1: Long papers)* (Vol. 1, pp. 250–259).

Davidson, T., Warmsley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. In *Proceedings of the International AAAI Conference on Web and Social Media.*

De Choudhury, M., Jhaver, S., Sugar, B., & Weber, I. (2016). Social media participation in an activist movement for racial equality. In *Proceedings of the international aaai conference on web and social media* (pp. 92–101).

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878–1881.

Kapidzic, S., & Herring, S. C. (2015). Race, gender, and self-presentation in teen profile photographs. *New Media & Society*, *17*(6), 958–976.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp.

55–60).

Muddiman, A., McGregor, S. C., & Stroud, N. J. (2018). (re)claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, *0*(0), 1-13.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015* (Tech. Rep.). Austin, TX: University of Texas at Austin.

Preoţiuc-Pietro, D., Liu, Y., Hopkins, D., & Ungar, L. (2017). Beyond binary labels: political ideology prediction of twitter users. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (Vol. 1, pp. 729–740).

Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication*, *67*(5), 803–826.