

Incivility Is Rising Among American Politicians on Twitter

Jeremy A. Frimer¹ , Harinder Aujla¹, Matthew Feinberg² ,
Linda J. Skitka³ , Karl Aquino⁴, Johannes C. Eichstaedt⁵,
and Robb Willer⁵

Social Psychological and
Personality Science
1–11
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/19485506221083811
journals.sagepub.com/home/spp


Abstract

We provide the first systematic investigation of trends in the incivility of American politicians on Twitter, a dominant platform for political communication in the United States. Applying a validated artificial intelligence classifier to all 1.3 million tweets made by members of Congress since 2009, we observe a 23% increase in incivility over a decade on Twitter. Further analyses suggest that the rise was partly driven by reinforcement learning in which politicians engaged in greater incivility following positive feedback. Uncivil tweets tended to receive more approval and attention, publicly indexed by large quantities of “likes” and “retweets” on the platform. Mediation and longitudinal analyses show that the greater this feedback for uncivil tweets, the more uncivil tweets were thereafter. We conclude by discussing how the structure of social media platforms might facilitate this incivility-reinforcing dynamic between politicians and their followers.

Keywords

incivility, political polarization, Twitter, social media, affective polarization

A thriving representative democracy requires that politicians adhere to a set of informal norms and rules of behavior that help temper antagonisms that can arise when people with differing views attempt to coexist and potentially work together (Levitsky & Ziblatt, 2018). Among the more commonly discussed norms is one proscribing incivility, which we define as impolite language that observers typically find to be rude and/or disrespectful (Mutz & Reeves, 2005).

Previous research has established that incivility harms individuals, relationships, and democracies. Being the target of, or merely a witness to, incivility has been shown to divert attention from critical issues (Brady & Crockett, 2019) and impedes creative, open-minded, and effective cognitive processing (Porath & Erez, 2009; Riskin et al., 2017). Incivility also tends to evoke anger and a desire to retaliate against the norm violator (Nugier et al., 2007; Wang & Silva, 2018). These reactions can impede the cooperation and joint problem-solving needed for a democracy to function (Anderson et al., 2014). Incivility may at times be of strategic value to politicians seeking to communicate moral clarity, signal their loyalty, fundraise, draw attention to issues, or mobilize their supporters to action (Brooks & Geer, 2007; Mason, 2018). When taken too far, however, political incivility can undermine respect for alternative viewpoints, erode public trust in the political process, and incite other forms of uncivil, undemocratic behavior (e.g., Mutz, 2015; see Lau et al., 2007, for a meta-analysis).

In recent years, scholars, journalists, and citizens have expressed concern about an apparent increase in the level of uncivil rhetoric in public discourse and in particular on social media (Haidt & Hetherington, 2012; Iyengar et al., 2012; Mann & Ornstein, 2012; Schaeffer, 2020; Shea & Fiorina, 2013). Previous work has focused on media (e.g., Sydnor, 2019) and citizens (e.g., Su et al., 2018) as purveyors of incivility, largely neglecting an important and influential source: elected officials themselves. One study (Theocharis et al., 2020) examined trends in elite political incivility over a 2-year period on Twitter and failed to detect any overall trend. We expand this investigation by establishing trends in levels of incivility over the span of a decade in 1.3 million tweets from the Twitter feeds of members of the U.S. Congress.

We focus on the language of elected officials on social media because politicians are tasked with striking compromises with their opponents for democracies to function

¹University of Winnipeg, Manitoba, Canada

²University of Toronto, Ontario, Canada

³University of Illinois at Chicago, USA

⁴University of British Columbia, Vancouver, Canada

⁵Stanford University, CA, USA

Corresponding Author:

Jeremy A. Frimer, Department of Psychology, University of Winnipeg, 515 Portage Avenue, Winnipeg, Manitoba, Canada R3B 2E9.
Email: j.frimer@uwinnipeg.ca

effectively. Incivility on the part of political elites can undermine this fundamental process in representative democracies. Furthermore, politicians shape the tone of discourse in the mass public, influencing how citizens communicate with one another on potentially divisive issues (Huddy & Yair, 2019). We focus on social media and Twitter in particular as it is among the primary modes by which U.S. politicians communicate with one another and the public (Pew Research Center, 2020; see the Discussion for an exploration of trends in incivility in other forums).

Prior research explains the rise of related forms of polarization with reference to a weakening of social norms prohibiting intergroup aggression (Iyengar & Westwood, 2015), the emergence of political action committees (“super PACs”; Haidt & Hetherington, 2012), partisan media (e.g., Fox News, MSNBC) and echo chambers (Rountree, 2019), and to an increase in negative political campaigning (Iyengar et al., 2012). We propose a novel explanation. Drawing from critical observations about the structure of social media, coupled with reinforcement learning principles (Skinner, 1948; Thorndike, 1898), we propose that political incivility might be rising on social media because of the operative social feedback mechanisms on the platform that incentivizes incivility.

Twitter limits and shapes the manner in which people interact. The most prevalent feedback mechanisms are the options to “reply,” “retweet” (share), or “like” a message, whereas other reactions, such as “disliking” are not on the menu of possible responses. The platform publicly displays, as metrics, tallies of how many times each of these occurred. Twitter users including politicians might attend to “reply,” “retweet,” and “like” counts as indicators of the social response to the original message.

The available scientific evidence appears to suggest that incivility is well received on Twitter. Uncivil and outrageous tweets tend to be rewarded with greater attention in the form of “retweets” (e.g., Brady et al., 2017) and we find that uncivil tweets also receive many “likes.” Politicians seeking to serve their constituents might interpret these metrics to mean that their followers approve of their messaging. We propose and present evidence that politicians learn from these social rewards and follow with further incivility in subsequent messages, yielding rising incivility on the platform.

The Present Study

We tested whether incivility is increasing over time in the Twitter messages of members of the U.S. Congress over the last decade and include presidential analyses for more descriptive purposes as social media has been prevalent for only two presidencies (Obama and Trump). Further analyses examine (and rule out) the possibility that an observed rise in incivility is fully attributable to several potential confounds such as uncivil politicians replacing civil ones in

Congress, rising negativity or moral outrage, rising polarization, or demographic shifts in Congress.

Prior research (Brady et al., 2017) found that uncivil and outrageous tweets tend to receive much publicity (i.e., “retweets”). We corroborate this finding and show that uncivil tweets are also rewarded with many “likes.” We then investigate whether the outsized reaction to incivility reinforces the original uncivil behavior and thus encourages more of it. Using a mediational framework, we test whether reactivity to uncivil tweets has also increased over time and whether statistically controlling for this rising reactivity reduces or eliminates the conditioned rate of increase of incivility over time. Noting the limitations of correlational inference, we augment this analysis with a longitudinal framework and a Granger causality analysis, testing whether intermittent periods in which uncivil rhetoric happened to receive an especially strong reaction on the platform were followed by a subsequent escalation of incivility within a Twitter feed.

Method

Samples

To test whether incivility is rising in the public statements of U.S. politicians, we acquired large corpora of tweets by members of the U.S. Congress and U.S. Presidents over the span of a decade. Sample size was determined by public availability: Our goal was to gather all publicly available materials. Our smallest sample (Presidential debates) had 80% power to detect effects of $|d| > 0.06$. We report all measures, conditions, and data exclusions, and how we determined sample sizes. This manuscript includes all studies attempted on this line of work.

Congressional Tweets

Although Twitter launched in 2006, we analyzed tweets beginning in 2009 because there were very few Congressional tweets prior. We, therefore, acquired the Twitter feeds of all members of the 111th–116th U.S. Congresses (2009–2019) and removed 350,039 “retweets,” leaving $N = 1,293,753$ original tweets in this sample. The first step to obtaining this sample was to gather the Twitter handles of all members of Congress. For the latest (116th) Congress in the sample, a Twitter list (<https://twitter.com/cspan/lists/members-of-congress/members>) provided the handles. For previous Congresses, we searched Twitter manually for each member and used their verified account or whichever was most evidently their professional Twitter account. For the 1,003 persons that served in the 111th–116th Congresses, we successfully gathered 841 handles (84%). In the 111th–116th Congresses, we identified 404, 458, 503, 513, 535, and 535 handles, respectively. With each tweet’s text, we collected the date and the number of “likes” and “retweets” it received.

To acquire the sample, we used a Python program (`tweets_sentiment_congress_heads.py`; see https://osf.io/thnuk/?view_only=d461532681bc49569d4a8222fe209644) to provide developer credentials via the Tweepy API (http://tweepy.readthedocs.io/en/v3.8.0/streaming_how_to.html) to collect tweet text, timestamps, favorites (i.e., “likes”), and “retweet” counts for each handle (at the time of data collection). We removed duplicate tweets. The Tweepy API and other interfaces for the Twitter API permits collecting up to 3,240 tweets per handle, which were sufficient to collect a complete set of tweets for most handles. Some handles had more than 3,240 tweets, meaning that the data set did not include all of their tweets. Moreover, politicians that were more active on Twitter (operationalized as the number of tweets collected per handle) were also more uncivil (operationalized as the average incivility across all of their tweets), $r(830) = .12, p < .001$. This introduces a potential confound when observing changes in incivility over time. We address this confound by including models that examine whether incivility increased over time within Twitter feeds.

In this (Congressional) and the Presidential Twitter samples, we excluded “retweets” for three reasons. First, “retweets” do not contain content written by the handle owner or his or her staff. Second, “retweeting” can sometimes but does not always signify an endorsement (see Study S4). And third, Twitter assigns “likes” of “retweets” to the original tweet, not to the “retweet” itself. All “retweets” “like” counts are identically zero and are therefore not amenable to testing our metric reinforcement explanation for the proposed rise in online incivility. We did not include analyses of “retweets with comments” because this feature was rare in this data set and unavailable before 2015.

Both the “likes” and “retweets” metrics had skewed distributions (see Table S1). To correct the skew, we applied log-transformations (after replacing 0s with the score 0.1, following Frimer, 2020) to retain all data and the ordinal nature.

Presidential Tweets. For descriptive purposes, we collected all 8,189 original tweets from Barack Obama’s Twitter handle (@barackobama) and 9,503 tweets from Donald Trump’s (@realDonaldTrump) while each of them was in office. Tweets from Barack Obama were collected using <https://github.com/twintproject/twint> for 2012–2017 and by manually acquiring tweets issued prior to 2012. Tweets from Donald Trump were downloaded from <http://www.trumptwitterarchive.com>. Tweets were restricted to the time that each President was in office and concluded on 21 November 2019. With each tweet’s text, we collected the date and the number of “likes” and “retweets” it received. Like in the Congressional Twitter sample, both metrics had skewed distributions. We corrected for skew in the same way as for the Congressional tweets.

Civilian Tweets. To assess whether changes over time in Congressional incivility might be attributable to a general rise in incivility on Twitter (i.e., among non-politicians too), we collected a pseudo-random sample of 998,935 tweets from the United States over the same time period, using approximately 100,000 tweets per year.

Measures of Incivility

Analyzing the level of incivility of each statement in such large corpora is infeasible using human judges. Therefore, we used four text analysis tools to quantify the level of incivility in texts. The primary tool was PerspectiveAPI’s “toxicity,” which we independently validated (see Study S1). PerspectiveAPI scores text for the level of incivility (referred to as “toxicity” in the PerspectiveAPI documentation) on a continuous scale from 0 to 100, reflecting the probability that an observer would find the tweet to be rude or disrespectful. It was designed to assist with moderating discussion forums (Adams, 2018) and trained using machine learning on hundreds of thousands of texts that had been annotated by thousands of human judges. (<https://cran.r-project.org/web/packages/peRerspective/peRerspective.pdf>). We independently validated the accuracy of PerspectiveAPI at judging levels of incivility (convergent validity with human observers $r = .627$ with $r = .647$ for Democratic observers and $r = .581$ for Republican observers; Study S1). Table 1 presents the examples of statements rated as low and high in incivility by the classifier in 2009 and 2019.

To code the incivility of each statement, we had Python code transmit one tweet/speech/debate statement per second to <http://perspectiveapi.com>. PerspectiveAPI can analyze any text up to 3000 characters in length. PerspectiveAPI toxicity scores were normally distributed (skews < 2), with sample M s ~ 13 and SD s ~ 10 (see the Supplement for details). To make results interpretable in real-world terms, incivility scores were retained in their raw form in the main analyses. In effect size analyses, incivility scores were standardized (z -scores).

Covariates

Any changes over time in the level of incivility in Congressional tweets are correlational and thus prone to third variable explanations. To establish whether changes in incivility over time are attributable to related or more general changes in language, such as rising negativity, we included measures of overall emotional tone and moral outrage. To assess whether the changes are attributable to a shifting political landscape, we included measures of politician partisanship and district safety. We included demographic variables to assess whether changes in incivility are attributable to shifting demographics (see the Supplement for details about these variables).

Table 1. Illustrative Examples of How the AI Classifier Characterized the Levels of Incivility in Tweets and Speeches by Members of the U.S. Congress in 2009 and 2019

Year	Mode	High incivility	Low incivility
2009	Tweet	@JoeSestak goes AWOL: Worst attendance out of all PA Reps, and 13th worst in all of Congress. <i>Sen. Arlen Specter (D-PA)</i> <i>AI-scored Incivility = 45.1</i>	The Stupak and Nelson amendments would disrupt the 30-yr status quo on abortion coverage to the disadvantage of womens' reproductive health. <i>Sen. Arlen Specter (D-PA)</i> <i>AI-scored Incivility = 6.8</i>
	Speech	The Pelosi health mandate bill . . . is going to destroy our economy . . . This is a dead, rotten, stinking fish that the Speaker is trying to force down the throats of the American people . . . <i>Rep. Keith Broun (R-GA)</i> <i>AI-scored Incivility = 46.0</i>	As House Republicans offer plans and ideas to get our economy moving again, all we get in return is more of the same, spending and taxing . . . we need new ideas . . . to deliver different results. <i>Rep. Adrian Smith (R-NE)</i> <i>AI-scored Incivility = 6.5</i>
2019	Tweet	Gov. Northam in Virginia is endorsing infanticide & proudly doing it! If that doesn't scare you, it should. It's horrifying . . . <i>Rep. Kevin Hern (R-OK)</i> <i>AI-scored incivility = 47.6</i>	. . . the total price tag of #GreenNewDeal would be . . . roughly four times the value of all Fortune 500 companies combined. That's no deal. <i>Sen. John Barrasso (R-WY)</i> <i>AI-scored Incivility = 6.4</i>
	Speech	This President and his Cabinet are so out of touch, it is pathetic . . . Shame on him and shame on Majority Leader McConnell for refusing to end this shutdown. He, along with the President, need to grow up. <i>Rep. Jim McGovern (D-MD)</i> <i>AI-scored incivility = 48.6</i>	I have heard stories of patients that are forced to choose between putting food on the table or paying for lifesaving drugs. Meanwhile, Big Pharma continues to collect hundreds of billions of dollars in profits each year. This cannot continue. <i>Rep. Debbie Mucarsel-Powell (D-FL)</i> <i>AI-scored incivility = 6.7</i>

Note. The AI classifier (Google's Perspective API) score texts on a scale from 0 to 100, reflecting the probability (0% to 100%) that a human observer would perceive the text to be rude or disrespectful. AI classifier scores converged with those from human ratings of perceived incivility, $r = .628$, and the accuracy was not conditioned by the political leaning of the politician and/or the judge (see Study S1).

Results

The 11-year analysis of 1.3 million tweets by members of the U.S. Congress revealed that levels of incivility rose on Twitter (Figure 1A). The test was a series of multilevel models, with time in years (mean centered at the year 2016.71) predicting incivility (PerspectiveAPI; 0–100 scale). In Model 1, we included no random effects. This model, thus, described the total observed change in incivility over time. In this model (see Table 2), incivility increased over time, representing a 23% rise from 11.8 in 2009 to 14.5 in 2019. Effect sizes (β s) were computed by standardizing the predictor and outcome variables and re-running the models. This total effect is likely to multiply determined, attributable to (a) politicians becoming more uncivil over time, (b) uncivil persons replacing less uncivil persons as members of Congress, and/or (c) particular features of the data collection method that included only more recent tweets from politicians.

Model 2 added random intercepts for each politician, effectively controlling for mean level differences between politicians. The observed increase of incivility over time in Model 2 is thus more likely attributable to politicians becoming more uncivil over time. Model 3 added random

intercepts for each politician and each session of Congress in a crossed random effects model, thereby controlling for the mean incivility of each cohort and modeling the average increase in incivility within each session, meaning over the 2-year period. Finally, Model 4 added random slopes for each politician and for each session, better isolating the change in incivility within each politician's Twitter feed. The effect of incivility over time decreased in magnitude slightly in the latter models but remained significant in all of them. Therefore, selection effects, whereby civil politicians drift away from Twitter and from Congress and uncivil politicians gravitate toward the platform and the chamber, might occur but cannot fully account for the observed increase. Politicians seem to have become more uncivil over time.

Incivility also rose in the general population over time, from 21.1 in 2009 to 27.0 in 2019, representing a 28% rise (recall that incivility rose by 23% among politicians over the same time span). An OLS regression confirmed that incivility rose among civilians, $B = 0.167$, 95% CI = [0.152, 0.183], $\beta = .022$, $p < .001$ (see Figure 1A). Interestingly, the rise in Congressional incivility was not explained by the general rise in the civilian population: A

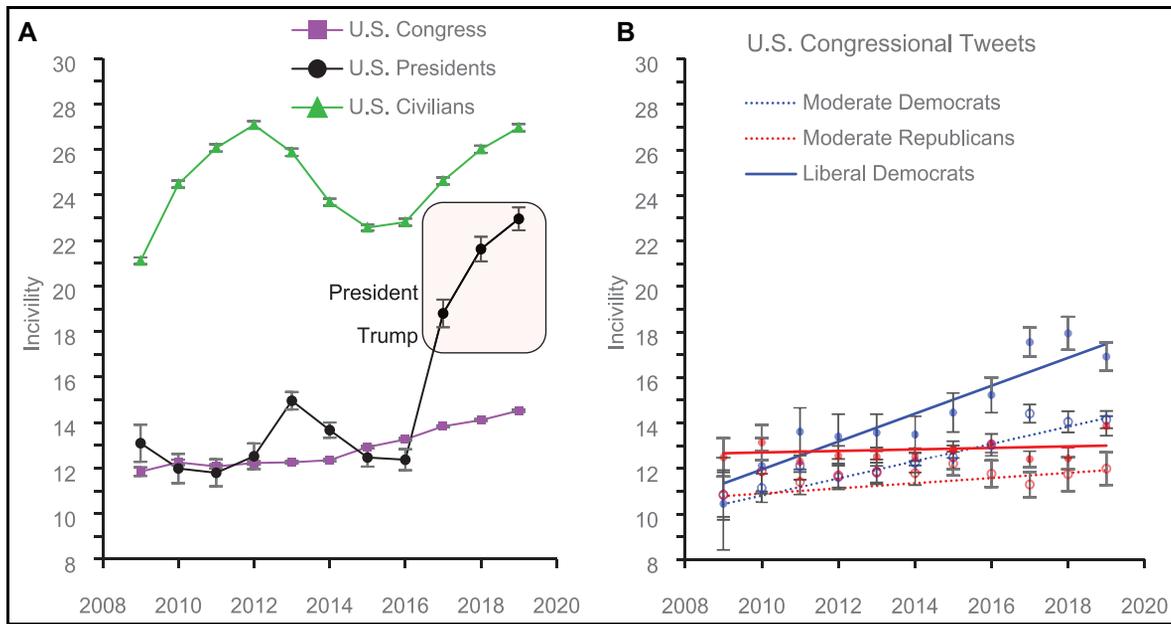


Figure 1. Changes in Levels of Incivility Among American Politicians Over Time on Twitter.

Note. Political incivility increased over time on Twitter both in the presidency and among members of the U.S. Congress (A), with the rise being the most pronounced among liberal Democrats (B). This partisan/ideological difference was mostly explained by reactions to President Trump (see the Supplement). Mean level estimates are from separate multilevel models for each year, with random intercepts for each Twitter user. Partisan levels (liberal, moderate, conservative) were derived from rollcall voting behavior. Trendlines were inferred from a party × partisanship × time multilevel model. Error bars represent 95% CIs.

Table 2. Changes in Levels of Incivility Among American Politicians Over Time on Twitter

Model	Random effects	Increase in incivility over time			
		B	95% CI	β	p
1. Tweets by politicians	(none)	0.332	0.325, 0.340	.076	<.001
2. Tweets by politicians	Intercepts for each politician	0.231	0.219, 0.242	.053	<.001
3. Tweets by politicians; politicians within sessions of Congress	Intercepts for each politician and for each session	0.252	0.231, 0.274	.058	<.001
4. Tweets by politicians; politicians within sessions of Congress	Slopes and intercepts for each politician and session	0.253	0.233, 0.275	.056	<.001

Note. The four multilevel models make different assumptions about the structure of the data and include different controls, but all point to the same conclusion that incivility increased over time. CI = confidence intervals.

multilevel model (Model 2 in Table 2) with time predicting Congressional incivility while controlling for the average civilian incivility in the year of each tweet found that incivility still independently rose among politicians over time, $B = 0.220$, 95% CI = [0.208, 0.232], $\beta = .051$, $p < .001$.

The finding of rising Congressional incivility on Twitter was generally robust across three other linguistic tools that capture elements of uncivil language: hate speech, uncivil words, and swear words (see Table S2). The rise in incivility generalized across the Senate and House of Representatives, as well as across the political spectrum (Figure 1B; see the Supplement for details).

Incivility scores in this study can range from 0 to 100. Scoring in the 10–25 range, the average tweets were not

particularly uncivil. Even the Twitter feed of President Trump produced scores that averaged around 20. This observation challenges the notion that most tweets by American politicians (including by President Trump) were uncivil. However, as we will later show, uncivil tweets receive much more attention than civil ones. The population might have had greater exposure to the relatively rare uncivil tweets seem more common than they actually were.

The rise in incivility over time held up when controlling for various potential confounds, each of which may have shifted over time and could have explained the rise. Our analytical approach was to estimate the rate of increase of incivility over time (in years) in a multilevel model (random intercepts for each politician) to establish an unconditioned

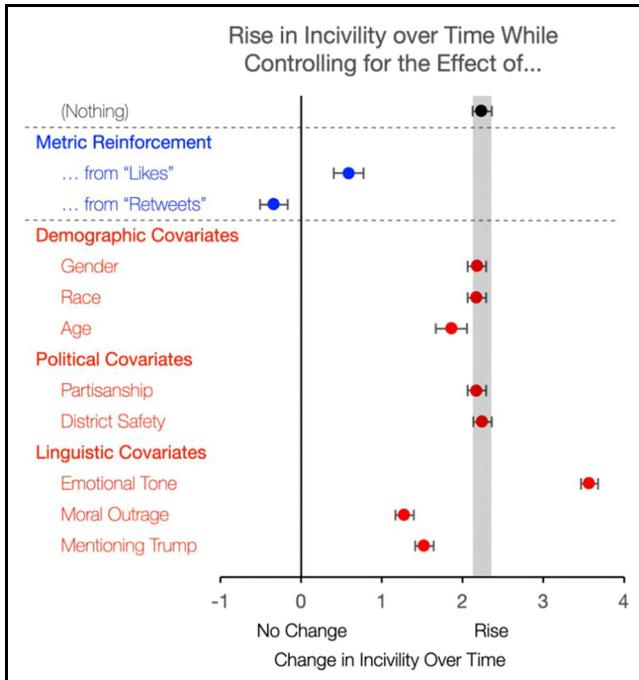


Figure 2. Conditioned Rise of Incivility While Controlling for Various Potential Explanations.

Note. The model was time predicting incivility while controlling for each potential mediator in separate analyses. The total effect of time on incivility is shown in the black dot with the colored dots being conditioned (indirect) effects of time on incivility. Statistically controlling for metric reinforcement (elevated “likes” and “retweets” from uncivil tweets) reduced or eliminated the observed rise in incivility over time (operationalized as PerspectiveAPI), meaning the metric reinforcement largely explained the rise in incivility. In contrast independently controlling for the effect of demographic, political, and linguistic covariates did not fully account for the observed rise in incivility over time. The increasing incivility over time is in units of change in incivility over 10 years with error bars being 95% confidence intervals.

baseline effect (Figure 2). We then added a covariate, one at a time and with no interaction terms, into a second multilevel model to test whether the conditioned effect of time on incivility remained. Evidence of mediation would be if the conditioned effect decreased or was eliminated.

The rise in incivility was not easily explained by parallel political, demographic, and rhetorical trends. We find that overall negativity in political tweets *decreased* over time, $B = -1.65$, 95% CI = $[-1.71, -1.60]$, $p < .001$, while incivility increased, meaning that incivility rose *in spite* of a tonal shift. Similarly, controlling for the effect of language communicating moral outrage (Brady et al., 2017) only partially accounted for rising incivility that we observed (Figure 2). A number of other potential confounds also stopped short of fully explaining the apparent rise. For example, the political polarization of Congress did not explain the rising incivility. Partisanship has risen in recent decades (Lewis et al., 2019), and partisans were found to be more uncivil than moderates (Figure 1). Yet controlling for partisanship did not account for the rise in incivility. The increase in incivility also remained when controlling

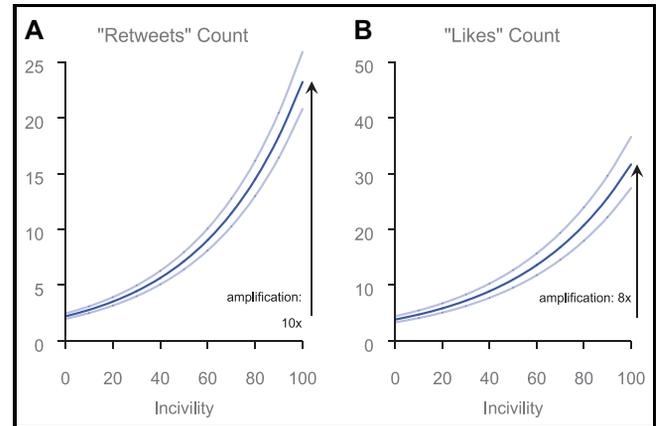


Figure 3. Consequences of Uncivil Tweeting.

Note. Uncivil tweets by members of the U.S. Congress received more “retweets” (A) and more “likes” (B) than civil ones. The distributions of “likes” and “retweets” were skewed. We performed log-transformations to normalize the data. Analyses were multilevel models, with random intercepts for each politician. We then removed the log transformation to plot on linear scales. Light lines represent 95% CIs.

for the diversification (along gender and racial lines) and aging of Congress, as well as political considerations (district safety). That said, rising incivility was partly (but not entirely) attributed to members of Congress addressing or talking about President Trump.

The age of Twitter only spanned the presidencies of Obama and Trump, limiting the scope of an analysis of incivility in the social media feeds of presidents. Incivility levels in President Obama’s Twitter feed did not change over time, averaging 13.4 (Figure 1A). When President Trump took office, levels of incivility in Presidential tweets jumped to 18.8 in 2017 and kept climbing to 23.0 in 2019, a 76% increase from 2009 to 2019, $d = 0.601$. The rise in online incivility during the Trump Presidency means that the rise in incivility in the Twitter feeds of U.S. presidents was not merely attributable to mean level differences between Presidents Obama and Trump.

Reinforcement Learning Explains the Rising Incivility

We found that uncivil tweets were apparently well received on Twitter, and this reception incentivized further incivility. Uncivil tweets received stronger social reactions in the form of many more “retweets” and more “likes” than civil tweets (Figure 3 and the Supplement for analyses). On average, extremely uncivil tweets (incivility = 100) received 10 times as many retweets and 8 times as many “likes” as extremely civil tweets (incivility = 0), $ps < .001$. If politicians are trying to be responsive to their constituents, it would be understandable if they were to escalate their incivility when the metrics to which they attend suggest that their incivility was well received.

Reactions to uncivil tweets compared with reactions to civil tweets became larger over time (Figure 4). In 2009,

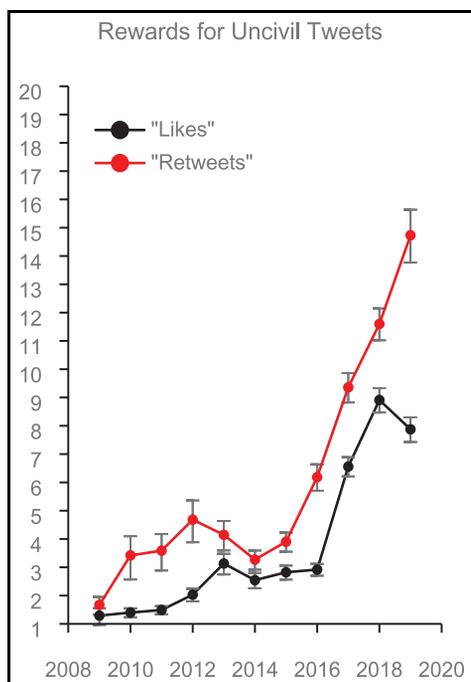


Figure 4. Rewards for Uncivil Tweeting Grew Stronger Over Time.

Note. The vertical axis representing the (model-implied) number of times more “retweets” and “likes” uncivil tweets (incivility = 100) received compared with civil tweets (incivility = 0). Error bars are 95% CIs.

uncivil tweets received approximately twice as many “retweets” as civil tweets, but by 2019, uncivil tweets received nearly 15 times as many. Similarly, uncivil tweets received 1.3 times as many “likes” as civil tweets in 2009 but nearly 8 times by 2019.

We tested whether these increasingly strong reactions to uncivil tweets accounted for the rise in incivility over time on Twitter using a mediational framework and found that they did (Figure 2). A conditioned effect of rising incivility over time was reduced by 74% when statistically controlling for the increasing reactivity in terms of “likes” and fully eliminated when controlling for reactivity from “retweets.” This finding stands in contrast to the limited mediational effects of controlling for various linguistic, political, and demographic factors.

This mediational analysis is correlational and thus prone to reverse causal and third variable interpretations. To more directly test whether reactions to incivility prompted more incivility from the politician, we leveraged the longitudinal nature of the data. Time can help tease apart correlational effects into their causal events because events that happen first can cause events that happen second but not vice versa. We, therefore, examined the reactions that uncivil tweets received were followed by a subsequent uptick in incivility on the part of the politician.

Uncivil tweets tended to receive more “likes” and “retweets” than civil ones (Figure 3). However, there existed variability within each Twitter feed in the degree to

which uncivil tweets provoked reactions. There were times for each member of Congress when their civil and uncivil tweets received similar reactions and other times when their incivility received a massive reaction compared with their civil tweets (see Figure S1). We exploited this variability within each Twitter account to test whether times when incivility was particularly well-received prompted a larger uptick in incivility than times when incivility was less well received (see Figure S2 for a conceptual illustration).

We formalized this notion with a Granger causality analysis, which allows for a form of causal inference (“Granger causality”) from longitudinal data. Formally, a Granger causal effect is established when there is an association between variable A at Time 1 and variable B at Time 2 while controlling for any autocorrelation in variable B at Time 1. In this case, variable A is the slope between the number of “likes” and the level of incivility in the 50 most recent tweets received. And variable B is levels of incivility in the most recent 50 tweets or the future 50 tweets (Equation 1; see the Supplement for a justification of a 50-tweet window and for similar results with 30- and 80-tweet windows; we did not impose date cut-offs):

$$Incivility_{future} = a \cdot Incivility_{past} + b \cdot \left(\frac{\# "Likes"}{Incivility} \right)_{past} + \varepsilon \quad (1)$$

$Incivility_{future}$ is the average level of incivility of the 50 next original tweets from the handle. $Incivility_{past}$ is the average level of incivility of the last 50 original tweets from the handle and serves as an autocorrelation control variable. And $\left(\frac{\# "Likes"}{Incivility} \right)_{past}$ is the slope of the number of “likes” and the incivility scores of the last 50 tweets from a handle and thus operationalizes reactivity to recent uncivil messages. An analogous formula was applied to reactions in the form of “retweets.” The parameter b is an indicator of whether learning has occurred and in what direction. If b is positive, the metric had a reinforcing effect on incivility. We fit these models for each of “likes” and “retweets” to assess which of the two metrics are implicated in learning.

The Granger causality analyses confirmed that politicians reacted to variability in reactions to incivility as theorized, $b_{retweets} = 4.34$, 95% confidence interval = [4.05, 4.63], $b_{likes} = 3.51$, 95% confidence interval = [3.20, 3.82], $ps < .001$ (see the Supplemental Materials for full details). Together, these results provide converging evidence that politicians learned from their positive experiences with uncivil tweeting and thus doubled down.

Discussion

We present the first robust evidence that incivility is rising among American politicians on Twitter, a rise that was not explained by a general rise in incivility in the civilian population. The rise appears to be attributable to more uncivil

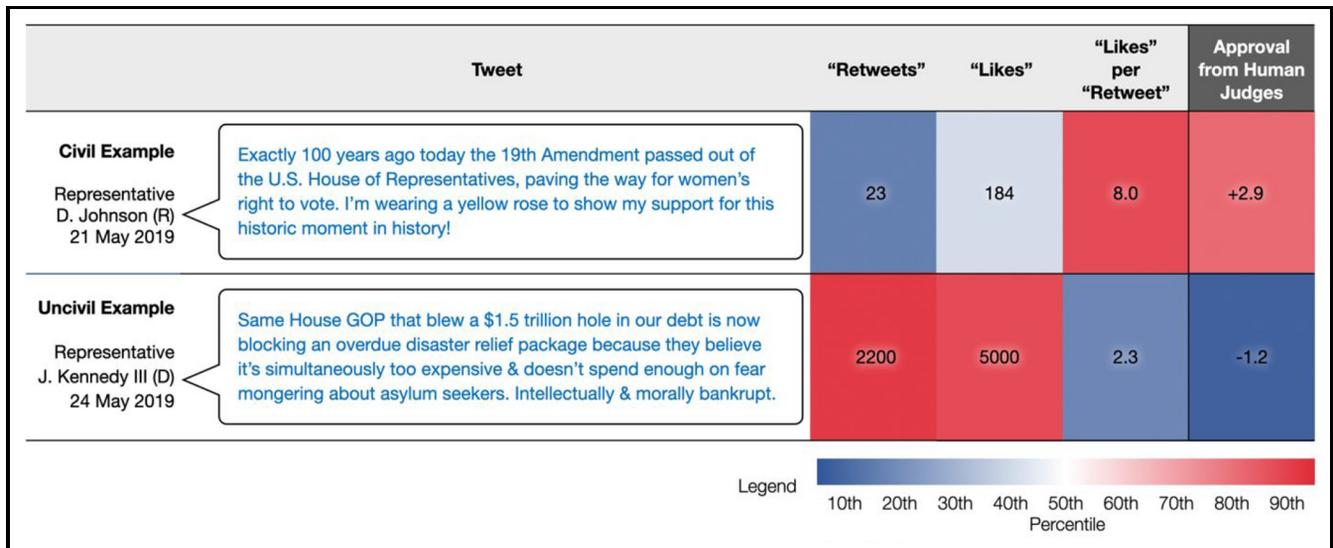


Figure 5. Examples of the distorting meaning of Twitter metrics.

Note. The top tweet by Rep. D. Johnson (R) is relatively civil and received relatively few “likes” or “retweet,” implying that it was not well received. Yet it received positive evaluations from an off-platform panel and a high “likes”-per-“retweet” ratio implying approval. In contrast, the bottom tweet by Rep. J. Kennedy III (D) is relatively uncivil, referring to political opponents as “intellectually & morally bankrupt.” It received many “likes” and “retweet” counts, implying a positive reception. Yet it received negative evaluations from an off-platform panel and a low “likes”-per-“retweet” ratio implying disapproval.

politicians replacing less uncivil politicians in Congress as well as to politicians themselves becoming more uncivil over time. Although the repercussions of this rise in incivility remain to be fully understood, the animosity and distrust it likely breeds are likely having some negative effects on American democracy and governance.

What is proximally causing the rising incivility? We offer the novel explanation that the rise is partly attributable to reinforcement learning on the platform. Uncivil tweets tended to receive many more “retweets” and “likes” than civil tweets. If politicians pay attention to these metrics and desire the attention and approval that they imply, they might learn that uncivil tweeting will help them politically. Both mediational and longitudinal analyses supplied supportive evidence of this explanation. The degree to which uncivil tweets received more “retweets” and “likes” than civil ones increased over the years under study; this historical shift mediated the rise in incivility over time. And using a longitudinal framework, we found that the Twitter community’s positive reactions to incivility predicted a subsequent uptick in incivility within the Twitter feeds of politicians.

If rising incivility on Twitter is partly due to platform-specific dynamics, as we find they might be, then we might expect to find that incivility has not risen in other contexts where social media dynamics are not in operation. Study S5 examined whether incivility rose in two offline contexts: Congressional speeches during floor debates between 1995 and 2020 and in Presidential debates between 1976 and 2016. We found that incivility did not rise and may have even been in decline in these in-person settings. These data

lend further credence to the notion that social media might uniquely foster incivility.

This juxtaposition between rising incivility on Twitter and flat or falling incivility in in-person exchanges raises questions about what, specifically, is the critical difference between these contexts that elicit different trends. One possibility is being and feeling physically distant from others on social media has a disinhibiting effect. Put simply, politicians might feel less inclined to be uncivil to a person’s face than they do in a private setting with their smart phone.

Another possibility is that features of the social media platform are playing an active role in fomenting incivility. By displaying quantities of reactions (e.g., “likes” counts), social media meaningfully altered the modes by which people provide social feedback to others. Jack Dorsey, the CEO of Twitter, acknowledged that the ramifications of these decisions have yet to be fully understood.¹ We propose that these metrics might not mean what they appear to mean, and that this distortion might be partly responsible for the rising incivility. On one hand, uncivil tweets tended to receive more “likes” than civil ones (Figure 3), implying social approval for incivility. On the other hand, recent research established that people tend to disapprove of uncivil messages (Frimer & Skitka, 2018, 2020). If observers tend to disapprove of uncivil tweets, why do uncivil tweets receive so many “likes”?

Supplemental studies S2 and S3 made the provocative discovery that the “likes” metric *diverges* with social approval and therefore signals the opposite of the underlying psychological reality. Paradoxically, we found that the more “likes” a tweet received, the less people actually liked

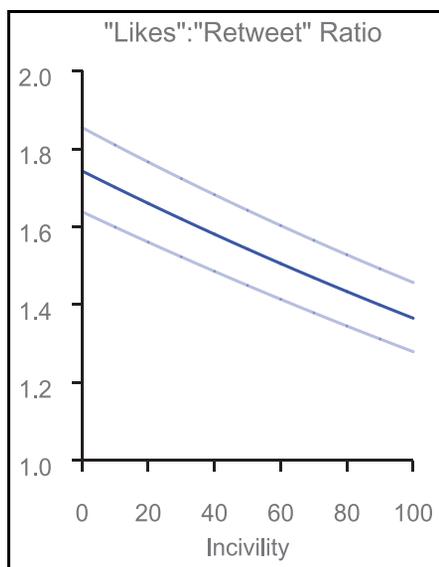


Figure 6. Social Disapproval of Uncivil Tweets.

Note. Uncivil tweets by members of the U.S. Congress had a lower “likes”-per-“retweet” ratio than civil tweets. This ratio tracks positively with levels of approval (see Studies S2-S3) meaning that the Twitter community, in general, did not approve of uncivil tweets in spite of appearances to the contrary. Light lines represent 95% CIs.

it (see Studies S2 and S3). Figure 5 illustrates the paradoxical divergence of Twitter metrics and social evaluations in two tweets by members of Congress.

The elevated “likes” counts that uncivil tweets received is likely attributable to their elevated prevalence. Uncivil tweets tend to have a larger audience than civil ones because uncivil tweets receive 10 times more “retweets” (Figure 3). As the typical Twitter user has 20–400 followers (Wojcik & Hughes, 2019), each “retweet” creates tens to hundreds of new impressions. Thus, uncivil tweets might be “liked” by a much smaller fraction of those who read them than civil tweets, but the uncivil tweets could still earn a higher absolute number of “likes” if many more people are exposed to those tweets. Although “retweeting” might also appear to signal social approval, Study S4 finds that the “retweeting” of uncivil tweets is driven by their entertainment value rather than by moral approval, a finding consistent with research showing that people feel compelled to pass on information about antisocial behavior without necessarily condoning it (Berger, 2011; Davis & McLeod, 2003; Feinberg et al., 2012). Together, these findings suggest that in spite of appearances the elevated “likes” and “retweets” the uncivil tweets tend to receive do not necessarily imply social approval.

To potentially correct for the confounding effect of exposure on the absolute count of “likes,” we created a “likes”-per-“retweet” ratio for each tweet and validated it as a valid indicator of social evaluation (Studies S2 and S3). Critically, uncivil tweets by members of Congress tended to have a lower “likes”-per-“retweet” ratio than civil tweets Figure 5, 6, implying that the Twitter community, including politicians’ like-

minded followers, tended to disapprove of uncivil tweets. This finding aligns with decades of research finding that political incivility draws disapproval (Lau et al., 2007), even from dedicated followers (Frimer & Skitka, 2018), and further suggests that metrics play a non-neutral role in breeding incivility.

Together, our results suggest that incivility is rising on Twitter in part due to platform-specific features, which are amenable to revision. Unless these platform-specific social dynamics are revised, it is likely that the rise in incivility that we observed among American politicians on Twitter will continue to rise. In light of the established deleterious effects that political incivility has on tolerance for alternative viewpoints and public trust in the political process (e.g., Mutz, 2015), this rise is likely to have negative effects on American democracy and governance.

Acknowledgments

The authors thank Pablo Barbera, Herbert Chang, Ari Decker-Frain, James Druckman, Emilio Ferrara, Amit Goldenberg, Justin Grimmer, John S. Hughes III, Austin van Loon, and Jamie Pennebaker for helpful comments, and Garrick Sherman for assistance with the civilian tweet sample.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: RW gave a speech at Twitter and received an honorarium. However, RW has no current professional relationships with Twitter and does not expect to receive any future compensation. All other authors declare no competing interests.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a grant from the Social Science and Humanities Research Council (SSHRC) of Canada to Jeremy A. Frimer (grant number 435-2018-1221).

ORCID iDs

Jeremy A. Frimer  <https://orcid.org/0000-0003-4811-0199>
 Matthew Feinberg  <https://orcid.org/0000-0001-8626-1620>
 Linda J. Skitka  <https://orcid.org/0000-0002-7078-0650>

Data and materials availability

Data and stimuli are available at <https://osf.io/thnuk/>. All data are available in the main text, the supplementary materials, or the online repository.

Data and materials availability

Data and stimuli are available at https://osf.io/thnuk/?view_only=d461532681bc49569d4a8222fe209644. All data are available in the main text, the supplementary materials, or the online repository.

Supplemental Material

The supplemental material is available in the online version of the article.

Note

1. In a 2020 interview, Jack Dorsey acknowledged that the ramifications of presenting quantities of social reactions might be complex and remained poorly understood, “The disciplines that we were lacking in the company in the early days that I wish we would have understood and then hired for are like a game theorist to just really understand the ramifications of tiny decisions that we make, such as . . . what happens when you put a count next to a “like” button” (Jackson, 2020)

References

- Adams, C. J. (2018, May 23). *New York Times: Using AI to host better conversations*. <https://www.blog.google/technology/ai/new-york-times-using-ai-host-better-conversations/>
- 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*, 373–387. <https://doi.org/10.1111/jcc4.12009>
- Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science, 22*(7), 891–893. <https://doi.org/10.1177/0956797611413294>
- Brady, W. J., & Crockett, M. J. (2019). How effective is online outrage? *Trends in Cognitive Sciences, 23*(2), 79–80.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences, 114*(28), 7313–7418. <https://doi.org/10.1073/pnas.1618923114>
- Brooks, J. B., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science, 51*, 1–16. <https://doi.org/10.1111/j.1540-5907.2007.00233.x>
- Davis, H., & McLeod, S. L. (2003). Why humans value sensational news: An evolutionary perspective. *Evolution and Human Behavior, 24*, 208–216. [https://doi.org/10.1016/S1090-5138\(03\)00012-6](https://doi.org/10.1016/S1090-5138(03)00012-6)
- Feinberg, M., Willer, R., Stellar, J. E., & Keltner, D. (2012). The virtues of gossip: Reputational information sharing as prosocial behavior. *Journal of Personality and Social Psychology, 102*, 1015–1030. <https://doi.org/10.1037/a0026650>
- Frimer, J. A. (2020). Do liberals and conservatives use different moral languages? Two replications and six extensions of Graham, Haidt, and Nosek’s (2009) moral text analysis. *Journal of Research in Personality, 84*, Article 103906. <https://doi.org/10.1016/j.jrp.2019.103906>
- Frimer, J. A., & Skitka, L. J. (2018). The Montagu Principle. Incivility decreases politicians’ public approval, even with their political base. *Journal of Personality and Social Psychology, 115*(5), 845–866. <https://doi.org/10.1037/pspi0000140>
- Haidt, J., & Hetherington, M. J. (2012, September 2017). Look how far we’ve come apart. *The New York Times*. <https://campaignstops.blogs.nytimes.com/2012/09/17/look-how-far-weve-come-apart/>
- Huddy, L., & Yair, O. (2019). *Reducing affective polarization: Warm group relations or policy compromise* [Paper presentation]. Paper presented at the annual meeting of the Western Political Science Association, San Diego, CA, United States.
- Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology: A social identity perspective on polarization. *Public Opinion Quarterly, 76*(3), 405–431. <https://doi.org/10.1093/poq/nfs038>
- Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party pines: New evidence on group polarization. *American Journal of Political Science, 59*, 690–707. <https://doi.org/10.1111/ajps.12152>
- Jackson, L. (2020). *Jack Dorsey on Twitter’s mistakes*. <https://www.nytimes.com/2020/08/07/podcasts/the-daily/Jack-dorsey-twitter-trump.html>
- Lau, R. R., Sigelman, L., & Rovner, I. B. (2007). The effects of negative political campaigns: A meta-analytic reassessment. *The Journal of Politics, 69*, 1176–1209. <https://doi.org/10.1111/j.1468-2508.2007.00618.x>
- Levitsky, S., & Ziblatt, D. (2018). *How democracies die* (1st ed.). Crown.
- Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., & Sonnet, L. (2019). *Voteview: Congressional roll-call votes database*. <https://voteview.com/>
- Mann, T. E., & Ornstein, N. J. (2012). *It’s even worse than it looks: How the American constitutional system collided with the new politics of extremism*. Basic Books.
- Mason, L. (2018). *Uncivil agreement*. University of Chicago Press.
- Mutz, D. C. (2015). *In-your-face politics. The consequences of uncivil media*. Princeton University Press.
- Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review, 99*(1), 1–15.
- Nugier, A., Niedenthal, P. M., Brauer, M., & Chekroun, P. (2007). Moral and angry emotions provoked by informal social control. *Cognition and Emotion, 21*, 1699–1720. <https://doi.org/10.1080/02699930601124738>
- Pew Research Center. (2020, July 16). *Congress soars to new heights on social media*. <https://www.pewresearch.org/internet/2020/07/16/congress-soars-to-new-heights-on-social-media/>
- Porath, C. L., & Erez, A. (2009). Overlooked but not untouched: How incivility reduces onlookers’ performance on routine and creative tasks. *Organizational Behavior and Human Decision Processes, 109*, 29–44. <https://doi.org/10.1016/j.obhdp.2009.01.003>
- Riskin, A., Erez, A., Foulk, T. A., Riskin-Geuz, K. S., Ziv, A., Sela, R., Pessach-Gelblum, L., & Bamberger, P. A. (2017). Incivility and medical team performance. *Pediatrics, 139*, Article e20162305. <https://doi.org/10.1542/peds.2016-2305>
- Rountree, C. (2019, March 9). *How American politics became to uncivil*. https://www.al.com/opinion/2016/03/how_american_politics_became_s.html
- Schaeffer, K. (2020). *Far more Americans see “very strong” partisan conflicts now than in the last two presidential elections*. <https://www.pewresearch.org/fact-tank/2020/03/04/far-more-americans-see-very-strong-partisan-conflicts-now-than-in-the-last-two-presidential-election-years/>
- Shea, D. M., & Fiorina, M. P. (2013). *Can we talk? The ride of rude, nasty, stubborn politics*. Pearson.
- Skinner, B. F. (1948). Superstition in the pigeon. *Journal of Experimental Psychology, 38*, 168–172.
- Sydnor, E. (2019). *Disrespectful democracy*. Columbia University Press.

- Su, Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Rude and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, 20(10), 3678–3699. <https://doi.org/10.1177/1461444818757205>
- Theocharis, Y., Barberá, P., Fazekas, Z., & Popa, S. A. (2020). The dynamics of political incivility on Twitter. *SAGE Open*, 10(2), 1–15. <https://doi.org/10.1177/2158244020919447>
- Thorndike, E. (1898). Some experiments on animal intelligence. *Science*, 8(18), 818–824.
- Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, 81, 73–83. <https://doi.org/10.1016/j.chb.2017.11.041>
- Wojcik, S., & Hughes, A. (2019, April 9). *Sizing up Twitter users*. <https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/>

Author Biographies

Jeremy Frimer is an Associate Professor of Psychology at the University of Winnipeg. His research focuses on politics and language.

Harinder Auja is an associate professor in the Department of Psychology at the University of Winnipeg. He is interested in computational approaches to understanding how experience influences semantic representation and in developing alternatives to traditional signal detection theory measures.

Matthew Feinberg is an Associate Professor of Organizational Behaviour and Human Resource Management at the Rotman School of Management at the University of Toronto. He explores the processes that unite people together to form cohesive and productive groups.

Linda Skitka is a LAS Distinguished Professor of Psychology and Political Science at the University of Illinois at Chicago. Her research is focused on understanding the cognitive and motivational underpinnings of ideological differences and the antecedents and consequences of moral conviction.

Karl Aquino is the Richard Poon Professor of Organizations and Society at the University of British Columbia. He conducts research on moral psychology, identity, victimization, and ideologically motivated thinking.

Johannes Eichstaedt is a Assistant Professor in Psychology and at the Institute of Human-Centered A.I. at Stanford. He focuses on mining large social media text data for public health and policy.

Robb Willer is a Professor of Sociology, Psychology, and Organizational Behavior at Stanford. His research focused on politics, morality, cooperation, and hierarchy.

Handling Editor: Will Gervais