

## Engagement With Social Media Posts in Experimental and Naturalistic Settings: How Do Message Incongruence and Incivility Influence Commenting?

XUDONG YU<sup>1</sup>

University of Amsterdam, The Netherlands

TERESA GIL-LÓPEZ

Universidad Carlos III de Madrid, Spain

CUIHUA SHEN

University of California, Davis, USA

MAGDALENA WOJCIESZAK

University of California, Davis, USA  
University of Amsterdam, The Netherlands

Research on factors that encourage people to speak out online uses either experimental or observational data, and it is unclear whether patterns detected in one setting apply to the other. This project examines whether message incongruence and incivility influence the likelihood of commenting on social media posts about politics in both controlled and naturalistic settings. An online experiment on American adults using a mock Facebook page (Study 1,  $N = 424$ ) showed that incivility decreased commenting when the original comment was pro-attitudinal but had no effects when it was counter-attitudinal and that incivility directly depressed commenting (i.e., unmediated via anger), but increased it indirectly through anger. An observational study on Spanish Twitter users (Study 2,  $N$  tweets = 4,153) demonstrated that in a naturalistic setting, there were more pro- than counter-attitudinal comments in response to the original tweet, and incivility was not associated with the desire to speak out. The implications are discussed.

*Keywords: political expression, message incongruence, incivility, social media, anger, third-person effect*

---

Xudong Yu: x.yu3@uva.nl

Teresa Gil-López: tgil@hum.uc3m.es

Cuihua Shen: cuishen@ucdavis.edu

Magdalena Wojcieszak: mwojcieszak@ucdavis.edu

Date submitted: 2021-06-27

<sup>1</sup> We thank Christopher Calabrese, Jingwen Zhang, and anonymous reviewers for their comments and suggestions.

Copyright © 2022 (Xudong Yu, Teresa Gil-López, Cuihua Shen, and Magdalena Wojcieszak). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

Political expression plays a vital role in the healthy functioning of a society. Expressing personal opinions can foment political participation (Vaccari et al., 2015) and increase political knowledge (Eveland, 2004), among other outcomes. Given these benefits, much research examines what influences individuals' willingness to speak out (e.g., Chung, Munno, & Moritz, 2015; Gearhart & Zhang, 2015; Ziegele, Weber, Quiring, & Breiner, 2018). One factor that is often tested is message incongruence (i.e., whether the stance advocated by a message is congruent with one's prior view). A message is congruent or pro-attitudinal if it supports one's existing political view; in contrast, a message is considered as incongruent, or counter-attitudinal, if it opposes one's prior viewpoint. For example, an article calling for stricter gun control is pro-attitudinal for individuals supporting gun control but counter-attitudinal to those who want to expand gun rights. Another factor that is frequently examined is incivility (i.e., whether the original post violates interpersonal politeness norms; Coe, Kenski, & Rains, 2014; Mutz, 2015; Mutz & Reeves, 2005).

Evidence on the role of these two factors in opinion expression online comes from either experimental or naturalistic settings, but never both. Individuals are known to behave differently in experimental settings, where they are captive participants "forced" to see certain political content, versus in naturalistic settings, where they use media at their own convenience and select content that they want to see or comment on (see Feldman, Stroud, Bimber, & Wojcieszak, 2013; Stroud, Feldman, Wojcieszak, & Bimber, 2019). As such, patterns found in one setting may not emerge in the other. To better understand human behavior and more accurately portray the factors influencing political expression online, it is important to analyze these factors in both settings. Despite key inherent differences in both settings, doing so also has unique and complementary advantages in terms of internal and external validity.

Furthermore, scant evidence systematically addresses how message incongruence and incivility influence self-expression on *social media* (but see Gearhart & Zhang, 2015), which are crucial avenues for discussion. About half of American adults use social media to receive news often or sometimes (Pew Research, 2021b) and about 4-in-10 users say social media are an important venue to express their political ideas (Pew Research, 2018).

To the best of our knowledge, this two-study project is the first to test the individual and joint effects of message incongruence and incivility on users' decision to comment on others' social media posts using both experimental and observational data. Study 1 is an online experiment with a sample of American adults ( $N = 424$ ). Study 2 uses behavioral trace data from Twitter ( $N$  tweets = 4,153). In both studies, we examine how (1) message incongruence, (2) incivility expressed in the original post, and (3) the interaction of message incongruence and incivility influence commenting. In Study 1, we take advantage of a controlled experimental design to test two theoretical mechanisms through which these two factors can encourage self-expression: third-person effect generated by message incongruence and anger stimulated by message incongruence and incivility.

### **Message Incongruence, Incivility, and Political Expression**

Among numerous factors that may affect the likelihood that citizens express their political opinions is message incongruence. It is important because if people mostly comment on pro-attitudinal content, they will be echoing and reinforcing their prior attitudes (Binder, Dalrymple, Brossard, & Scheufele, 2009); and if they comment on counter-attitudinal posts, they may start deliberative conversations (Price, Cappella, & Nir, 2002) or conflicts (Mendelberg & Oleske, 2000).

Another factor influencing one's desire to speak out on social media is incivility, a concept that is hard to define "because what strikes one person as uncivil might strike another person as perfectly appropriate" (Coe et al., 2014, p. 660). Attempts to define incivility can be summarized into two broad categories. Some researchers claim that incivility is the violation of interpersonal politeness norms (e.g., "This clueless GUNNUT proves my point"; Coe et al., 2014; Mutz, 2015; Mutz & Reeves, 2005). Under this definition, incivility is operationalized as the disrespectful tone of political discourse, usually encompassing name-calling, aspersions, and other impolite behavior (e.g., Coe et al., 2014; Mutz, 2015; Rowe, 2015).

Others argue that incivility is disrespect for democratic ideals, and a set of behaviors is uncivil only when they "threaten democracy, deny people their personal freedoms, and stereotype social groups" (e.g., "Those who do not have a college degree should not be allowed to vote"; Papacharissi, 2004, p. 267). Accordingly, messages are uncivil if they harm or discriminate against other people or groups (Rossini, 2020), such as offensive stereotyping, violent threats, and other antidemocratic expressions (Rossini, 2020; Rowe, 2015). Reconciling the two different approaches, yet other scholars assert that incivility is a two-dimensional term, labeling the former as impoliteness or personal-level incivility and the latter as intolerance or public-level incivility (Muddiman, 2017; Rossini, 2020; Rowe, 2015). In reality, personal-level incivility is more common than public-level incivility (Papacharissi, 2004; Rowe, 2015) and is perceived as more uncivil (Muddiman, 2017). In addition, most prior studies on online discussions focus on personal-level incivility (Rossini, 2020). Thus, this project examines the effects of personal-level incivility on commenting.

Estimates suggest that a substantial proportion of user-generated content online contains incivility. For instance, about 20% of tweets mentioning U.S. legislators were uncivil (Theocharis, Barberá, Fazekas, & Popa, 2020) and about 40% of comments on national as well as local news outlets' Facebook pages were classified as rude or *extremely* uncivil (Su et al., 2018).

Previous research has used experiments and observational studies to test the effects of message incongruence and incivility on self-expressions, but the patterns found from different methods are inconsistent. For example, using an experiment, Chung et al. (2015) showed that incongruent content predicted speaking out; however, relying on Twitter data, Williams, McMurray, Kurz, and Lambert (2015) demonstrated that people were more likely to "mention" those who shared similar views in their messages, indicating that congeniality enhances online expression. In terms of incivility, experimental data showed that uncivil messages increased the willingness to reply (Ziegele et al., 2018) while an observational study found that offensive comments got only slightly more feedback (Ziegele, Breiner, & Quiring, 2014).

These discrepant findings may be because of different information environments. In experiments, researchers create artificial information spaces where participants are usually equally likely to encounter certain information types (e.g., congenial or cross-cutting, civil or uncivil; e.g., Chung et al., 2015). In reality, however, individuals prefer pro-attitudinal (e.g., Garrett, 2009) and civil messages (GU Politics, 2019; Montanaro, 2018), making such experimentally “forced” exposure unlikely to occur in people’s daily lives. Therefore, to offer a more general—and likely more accurate—understanding of how these two factors promote commenting on social media posts, we examine the effects of message incongruence and incivility in both experimental and naturalistic settings.

### Experimental Settings

We propose that, in experiments where people are equally likely to see pro- and counter-attitudinal comments, individuals are more likely to comment on the latter because dissimilar information can evoke third-person effect and anger, both of which predict a stronger desire to speak out. According to the third-person effect theory (TPE; Davison, 1983), people tend to think that media content has a larger effect on other people than themselves, and this effect is much stronger for messages with harmful outcomes (e.g., Gunther, 1995). Because counter-attitudinal messages can be seen as harmful or misleading by users, these messages should be especially likely to evoke the TPE. As further predicted by the corrective action hypothesis, individuals may want to take action when they perceive TPE, such that “people would engage in reactive actions to have their own views be heard in order to counterbalance those perceived media effects” (Rojas, 2010, p. 347). Commenting on counter-attitudinal posts can be an effective and direct way to counteract its perceived negative effects. In contrast, positive information (such as pro-attitudinal messages), either evokes no TPE (Gunther & Mundy, 1993) or leads people to think that they are more impacted by such information than others (David, Liu, & Myser, 2004), so its effects on commenting should be limited.

Additionally, counter-attitudinal posts can stimulate anger, an emotion that arises when people feel offended or when they are kept from reaching their goals (Nabi, 2002). When reading such posts, individuals may believe that their own positions are threatened, which then evokes anger. Indeed, Arpan and Nabi (2011) found that counter-attitudinal messages cause anger, which, in turn, predicts the desire for retaliation (see also Nabi, 2002). In the case of social media, replying to a post that challenges individual prior beliefs is an easy and feasible “proxy” for such a revenge. Accordingly, research showed that individuals supporting gun rights were more likely to report intentions to comment on news after reading pro-gun control stories than stories supporting gun rights, and vice versa for gun control supporters (Chung et al., 2015). We measure actual commenting *behaviors*, rather than self-reported intentions, while also accounting for posts’ incivility.

Concerning incivility, in experimental settings where the chances of encountering civil and uncivil posts are equal, users are more likely to comment on the latter because incivility induces anger. Incivility breaks social norms (Andersson & Pearson, 1999; Mutz, 2015), which produces anger (Ohbuchi et al., 2004). For instance, people who witnessed incivility in a work setting reported feeling angry (Porath, MacInnis, & Folkes, 2010) and Gervais (2017) found that exposure to uncivil political messages increased anger. As aforementioned, anger predicts motivations to take action, thereby leading to greater willingness

to comment. Indeed, Ziegele et al. (2018) found that participants who saw an uncivil comment on a news article reported greater willingness to reply compared with those who read a civil comment.

How do message incongruence and incivility interact? We speculate that incivility will attenuate the willingness to comment on pro-attitudinal posts. Such messages do not generate anger (Gervais, 2015, 2017) because individuals are not the target of incivility (Gervais, 2017; Phillips & Smith, 2004). Instead, people may feel sad or anxious and attempt to distance themselves from uncivil ingroup members (Druckman, Gubitz, Lloyd, & Levendusky, 2019), thereby avoiding interacting with them online. In contrast, incivility should encourage commenting on counter-attitudinal posts. Relative to counter-attitudinal *civil* posts, parallel *uncivil* ones evoke greater anger (Gervais, 2017) because they threaten individuals' prior beliefs *and* use inappropriate language, thus encouraging users to reply. Thus, we predict:

*H1a: In an experimental setting, people are more likely to comment on counter-attitudinal posts than pro-attitudinal posts.*

*H2a: In an experimental setting, people are more likely to comment on uncivil posts than on civil posts.*

*H3a: In an experimental setting, the effect of incivility on commenting is moderated by message incongruence, such that incivility reduces commenting when the posts are pro-attitudinal and promotes commenting when the posts are counter-attitudinal.*

*H4: The effect of message incongruence on commenting is mediated by (a) third-person effect and (b) anger, such that counter-attitudinal posts stimulate third-person effect and anger, which in turn encourage commenting.<sup>2</sup>*

*H5: The effect of incivility on commenting is mediated by anger, such that uncivil posts elicit anger, which in turn promotes commenting.*

*H6: The mediation effect of incivility on commenting through anger is moderated by message incongruence, such that the path through anger should be strongest for counter-attitudinal uncivil posts.*

### **Naturalistic Settings**

The above predictions assume that people are equally likely to encounter different types of posts, a situation that is unlikely to occur during actual online exposures and interactions. In such contexts, nonrandom circumstances influence not only the likelihood that individuals encounter specific social media contents, but also the ways in which they subsequently engage with them.

In a naturalistic setting, such as on an actual Twitter newsfeed, users can easily see more pro-attitudinal content. Factors such as network homophily reinforce selective exposure, as users tend to affiliate

---

<sup>2</sup> See details about how third-person effect and anger were measured in the measures section of Study 1.

with like-minded others (see e.g., Bakshy, Messing, & Adamic, 2015; Yarchi, Baden, & Kligler-Vilenchik, 2021). Networks then act as information disseminators and filters that further reduce the spectrum of content the user encounters. Users can also filter out undesirable contents by hiding or blocking posts or other users (e.g., John & Dvir-Gvirsman, 2015; Yang, Barnidge, & Rojas, 2017). Finally, algorithms generate content recommendations (e.g., "you may also like . . ."; Pariser, 2011). These three mechanisms (i.e., one's preference for like-minded content, network homophily, and algorithms) are likely to reduce users' exposure to counter-attitudinal posts. Engagement with pro-attitudinal content, therefore, seems more likely by the sole virtue of greater number of pro-attitudinal messages, even though users may in fact be inclined to comment on counter-attitudinal messages whenever they encounter them.

Still other aspects may determine the availability of congruent positions to users in the online realm. One of them is that social media platforms showcase distinct discursive affordances; their architecture shapes the way people interact. On Twitter, a platform we analyze, conversations are based on hashtags, which often connect users with similar opinions when hashtags denote preferred framings of an issue (Bruns & Burgess, 2011). On the other hand, hashtags enable "momentary connectedness" with broader, rather imprecise audiences (Rathnayake & Suthers, 2018, p. 2). In creating ad hoc discussion networks around specific topics, hashtags can effectively expand the boundaries of online discussions beyond the users' immediate network of self-selected connections. Specific political contexts may also influence distances among dissimilar viewpoints, thus further shaping the probability that users encounter counter-attitudinal messages; for instance, some debates tend to polarize positions (e.g., Balcells & Padró-Solanet, 2020, on Catalan independence; Evolvi, 2019, on anti- and pro-muslim narratives after Brexit).

Concerning incivility, past work suggests that users are less likely to encounter uncivil than civil posts in naturalistic settings because, although incivility is prevalent on social media, most content remains civil (Rowe, 2015; Su et al., 2018). In addition, because most citizens have negative attitudes toward political incivility (e.g., GU Politics, 2019; Montanaro, 2018), users are likely to avoid it by unfollowing, reporting, blocking, and/or unfriending those whose posts lack civility and respect.

It follows that, because pro-attitudinal and civil messages are more prevalent on users' social media feeds than other types of posts, users should engage with pro-attitudinal civil posts at greater rates. We advance three new hypotheses:

- H1b: In a naturalistic setting, there will be a higher proportion of pro-attitudinal than counter-attitudinal comments in response to a post.*
- H2b: In a naturalistic setting, there will be a higher proportion of comments on civil posts than on uncivil posts.*
- H3b: In a naturalistic setting, there will be a higher proportion of pro-attitudinal comments in response to civil than to uncivil posts.*

### **Data Limitations**

Before describing the designs, one core limitation should be acknowledged. The two studies do not only differ in their setting (experimental vs. observational) but also in terms of the country context (the United States versus Spain), the issue studied (gun control versus Catalan independence), and the platform (professionally designed mock Facebook page vs. Twitter). Thus, the different effects detected may not be fully because of experimental or observational settings but due to other differences. This key limitation should be kept in mind, and we acknowledge that unique assumptions operate in each setting. Nevertheless, both cases share a basic principle. Independent of how likely individuals are to encounter different posts in online spaces, there is arguably a minimum array of considerations and choices that they must ponder whenever they come across a social media post and face the decision of engaging with it: among others, whether or not they agree with their interlocutor, and/or if people should abide by certain norms of politeness. Hence, we believe the advantages provided by comparing the effects of message incongruence and incivility in different settings in a single article, especially given the dearth of such research, justify the effort. The two studies investigated the same phenomenon and tested closely related hypotheses, complementing each other. Specifically, Study 1 tested how message incongruence and incivility influence commenting in an idealized, artificial setting, and Study 2 was concerned with whether the patterns in Study 1 could also be found in the real world. Thus, presenting the two studies in one article provides a comprehensive picture of commenting behavior on social media.

### **Study 1**

#### **Method**

We first investigated the effects of message incongruence and incivility on commenting in an online experiment. Data were collected in July and August 2018 by Survey Sampling International. The sample included American Facebook users who were 18 or older. Because the study tested engagement with pro-versus counter-attitudinal posts on a contentious issue (i.e., gun control), participants who were truly neutral on the issue were automatically filtered out.

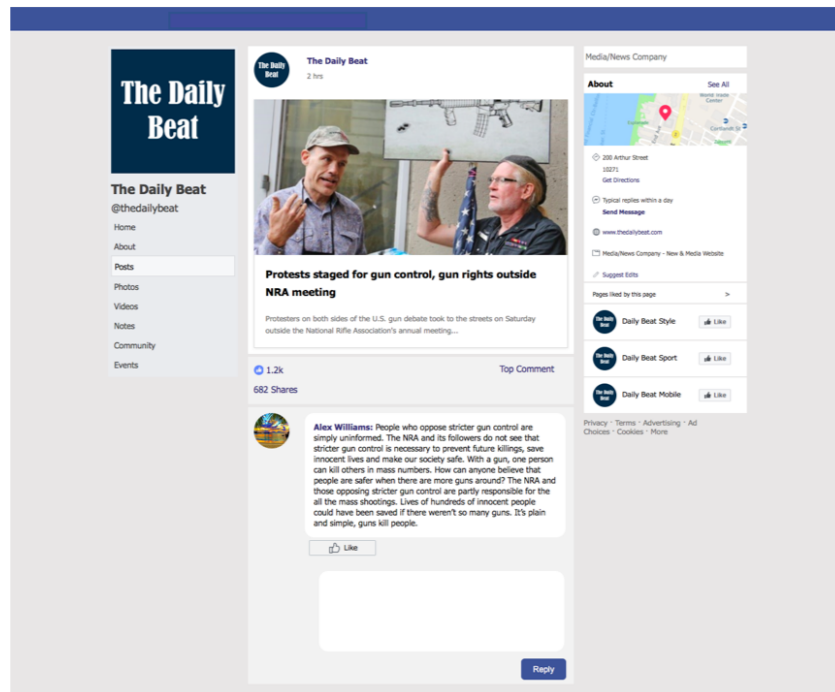
In total, 643 individuals completed the study. From this group, we excluded those who were speeders (48% of the median time as recommended by SSI) and who failed an attention-check question (see Appendix A for details).<sup>3</sup> The final sample included 424 participants (see the demographics in Appendix A).

After a pretest, participants were redirected to a mock Facebook page of a generic news organization ("The Daily Beat"). The page was professionally designed to resemble an actual Facebook experience (see Figure 1). Participants could click to read the article, read the comment under the post, reply to it, "like" (like, love, haha, wow, or yay) or "dislike" (angry, sad, or confused face) it, or do nothing

---

<sup>3</sup> Appendix available here: [https://drive.google.com/file/d/1x\\_\\_bS9yc0-apWut1pX\\_UAIfaX1oeIHxB/view?usp=sharing](https://drive.google.com/file/d/1x__bS9yc0-apWut1pX_UAIfaX1oeIHxB/view?usp=sharing)

at all<sup>4</sup> (other functionalities were disabled). The page presented a news article titled, “Protests staged for gun control, gun rights outside NRA meeting” (see text in Appendix B).



**Figure 1. Screenshot of the mock Facebook page.**

In the treatment conditions, we systematically varied the comment under the article, posted by a fictional Facebook user (a gender-neutral username Alex Williams and a generic picture that did not identify the gender, race, or age of the user; see Figure 1; see comments’ text in Appendix B). Participants were randomly assigned to one of four conditions: (1) civil pro-attitudinal ( $n = 109$ ), (2) uncivil pro-attitudinal ( $n = 99$ ), (3) civil counter-attitudinal ( $n = 112$ ), and (4) uncivil counter-attitudinal ( $n = 104$ ). Comment (in)congruence was determined based on one’s attitude toward gun control, assessed at the pretest. For instance, a participant who was favorable toward gun control saw a pro-gun control comment in the pro-attitudinal condition. Incivility was manipulated by name-calling, aspersion, and other uncivil elements (Coe et al., 2014) in the comment. The length (word count between 100 and 104) and content of the comment were held constant within the pro- and counter-attitudinal condition, presenting the same facts and arguments in either civil or uncivil way. Both the article as well as the comments had been pilot-tested to

<sup>4</sup> We did not explicitly tell participants that they could or should comment. We think it is unlikely that they did not realize they could if they wanted to. First, because we asked participants to do what they would normally do on Facebook. Second, because there was a comment box and a reply button under the post, indicating that they could comment (see Figure 1).



ensure that the article was perceived as neutral and the comments were perceived to align with their respective experimental conditions (see Appendix C).

Participants spent at least 90 seconds on the mock Facebook page before moving forward. After 90 seconds elapsed, they were presented with a button to the posttest, but they could continue browsing and engaging with the page for as long as they desired.

### Measures

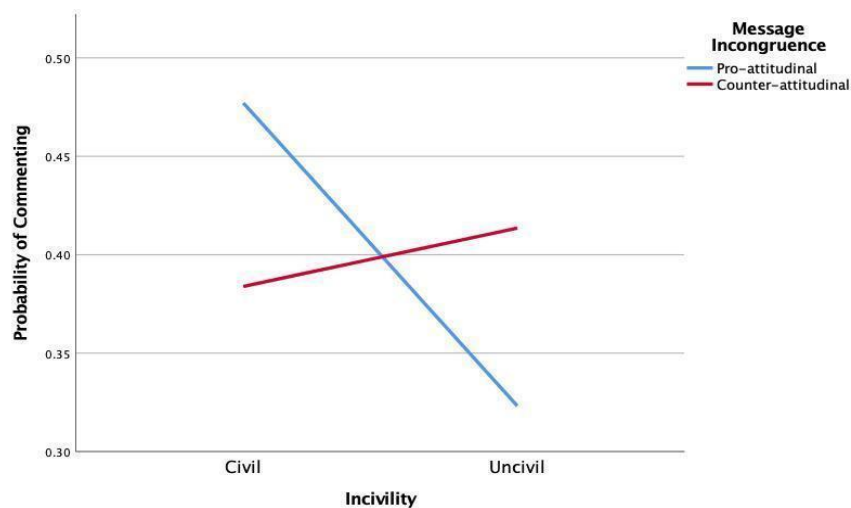
#### Behavioral Measures

The web server hosting the mock Facebook page recorded whether participants commented on the original comment. In total, 170 participants commented (40.09% of the sample).

#### Mediating Measures

At the posttest, participants indicated how much the original comment they saw made them feel (1) irritated and (2) angry (1 = "not at all," 7 = "very much"). These items were averaged ( $r = .88$ ;  $M = 3.54$ ,  $SD = 2.22$ ). Participants also reported how much influence they thought the comment might have on their own opinions toward gun control ( $M = 2.85$ ,  $SD = 2.18$ ) and on the opinions of others ( $M = 3.83$ ,  $SD = 1.77$ ; 1 = "no influence at all," 7 = "a lot of influence"). The final third-person effect measure was computed by subtracting the perceived effects of the comment on the self from the perceived effects on others ( $M = 0.98$ ,  $SD = 1.81$ ; Rojas, 2010).

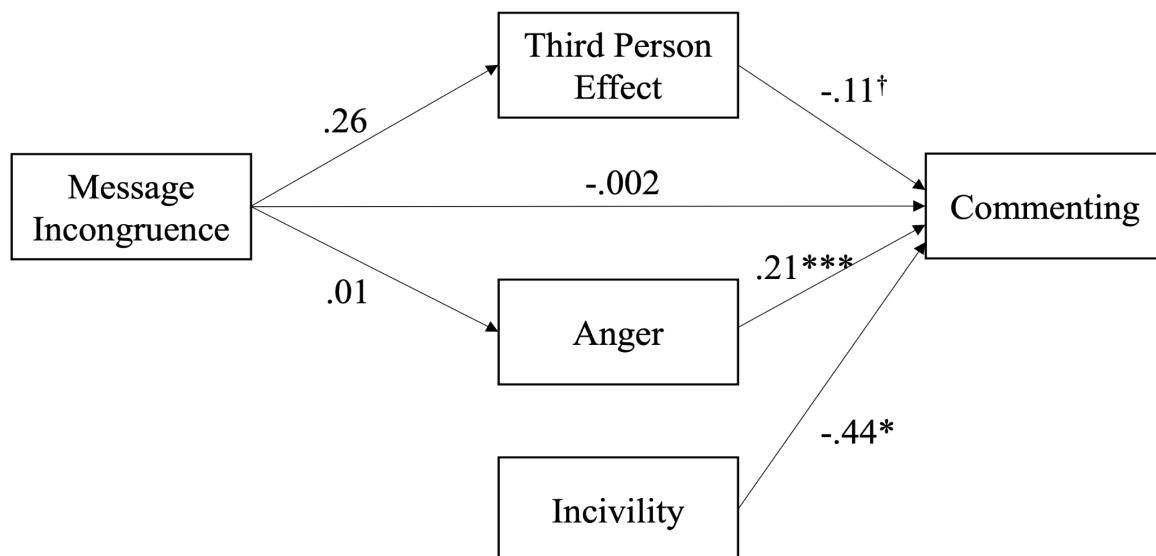
### Results



**Figure 2. Moderating effects of message incongruence on the relationship between incivility and commenting.**

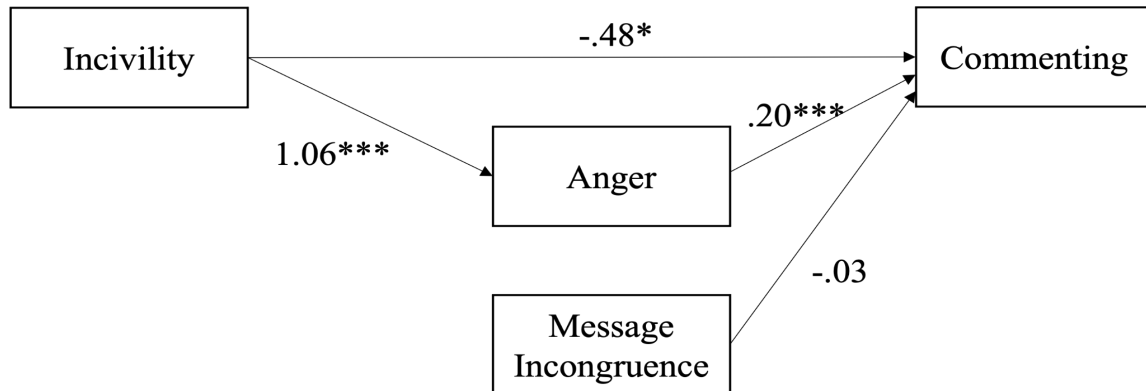
To test H1a and H2a, we estimated a logistic regression model with the recorded commenting entered as the dependent variable (did not comment = 0, commented = 1) and message incongruence and incivility entered as predictors, meaning that when examining the effects of one, we controlled for the other. Neither message incongruence ( $b = -0.02, p = .910$ ) nor incivility ( $b = -0.25, p = .205$ ) had statistically significant main effects on commenting, rejecting H1a and H2a. To test whether these factors interact (H3a), an interaction term was included in the model. The coefficient was positive but insignificant ( $b = 0.77, p = .054$ ), and the direction of the interaction can be seen from Figure 2. Incivility decreased the probability of commenting from .48 to .32 when the original comment was pro-attitudinal (simple effects coefficient:  $b = -0.65, p = .024$ ) but had no statistically distinguishable effects on commenting when the original comment was counter-attitudinal (simple effects coefficient:  $b = 0.12, p = .658$ ), offering partial support for H3a.

To test whether counter-attitudinal comments promote commenting through TPE and anger (H4), we employed regression-based path-analytic mediation tests, using bootstrapped confidence intervals estimated with 10,000 samples (PROCESS, Model 4, Hayes, 2017). The models tested TPE and anger as mediators with incivility entered as a covariate (see Figure 3). Contrary to the predictions, relative to pro-attitudinal comments, counter-attitudinal comment stimulated neither TPE ( $b = 0.26, p = .140$ ) nor anger ( $b = 0.01, p = .956$ ). Anger was positively associated with commenting ( $b = 0.21, p < .001$ ), whereas the association between TPE and commenting was not statistically significant ( $b = -0.11, p = .070$ ), and the indirect effects were nonsignificant (anger:  $b = 0.003, SE = .05, 95\% CI [-0.09, 0.10]$ ; third-person effect:  $b = -0.03, SE = .03, 95\% CI [-0.09, 0.01]$ ). Thus, both H4a and H4b are rejected.



**Figure 3. Mediation model including unstandardized regression coefficients for message incongruence as the independent variable, third-person effect and anger as mediators, and commenting as the dependent variable. Incivility was included as a covariate.  $^\dagger p < .10$   $^* p < .05$ ,  $^{**} p < .01$ ,  $^{***} p < .001$ .**

A parallel model examined whether incivility increased commenting through anger (H5), with message incongruence entered as a covariate (see Figure 4). Participants who saw uncivil comments reported greater anger compared with those who saw civil comments ( $b = 1.06, p < .001$ ), which predicted greater likelihood of commenting ( $b = 0.20, p < .001$ ), and a significant indirect effect was observed ( $b = 0.21, SE = .07, 95\% CI [0.09, 0.36]$ ), supporting H5. Interestingly, incivility directly depressed commenting (i.e., unmediated by anger;  $b = -0.48, p = .024$ ).



**Figure 4. Mediation model including unstandardized regression coefficients for incivility as the independent variable, anger as the mediator, and commenting as the dependent variable. Message incongruence was included as a covariate. †  $p < .10$  \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$ .**

H6 predicted that the mediation effect through anger will be strongest for counter-attitudinal uncivil posts. Incivility was entered in PROCESS model 7 as the independent variable with anger as the mediator, message incongruence as the moderator, and TPE as the covariate. Although anger was associated with greater commenting ( $b = 0.21, p < .001$ ), the interaction between message incongruence and incivility did not significantly influence anger ( $b = -0.49, p = .243$ ), leading to insignificant moderated mediation effects ( $b = -0.10, SE = .09, 95\% CI [-0.29, 0.08]$ ). Thus, H6 is rejected.<sup>5</sup>

### Discussion

Study 1 experimentally tested the influence of message incongruence and incivility on users' commenting on Facebook posts. Participants who were exposed to *civil* pro-attitudinal posts were more likely to comment than those encountered *uncivil* pro-attitudinal posts; however, users were equally likely to comment on civil and uncivil counter-attitudinal posts. Study 1 also showed that exposure to uncivil posts elicited anger, which led to greater likelihood of commenting. However, incivility, at the same time, directly decreased commenting (i.e., unmediated by anger).

<sup>5</sup> Models with education level entered as a covariate produced similar results.

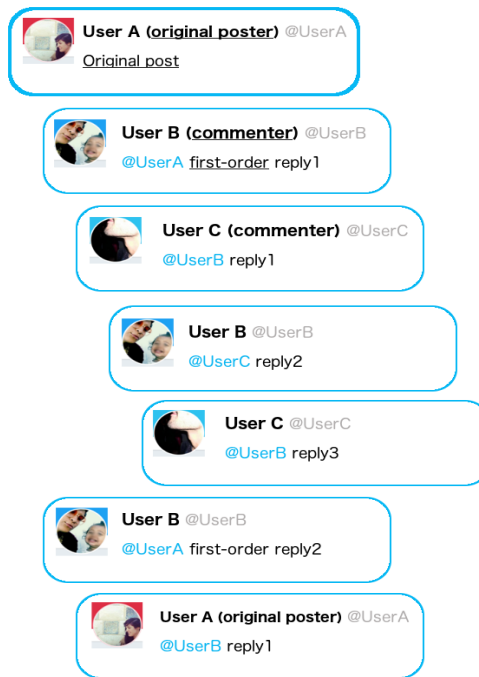
## Study 2

To examine commenting in a naturalistic setting, we used a Twitter data set of 735 protest-related discussion threads by 681 accounts between the 15 days preceding and the 15 following November 9, 2014, the day the independence referendum was carried out in the northeastern region of Catalonia, in Spain.<sup>6</sup> Although we draw on a distinct context, the debate over the Catalan independence in Spain displays levels of polarity akin to those over gun control in the United States. This enables a relatively clean distinction of groups in terms of their pro- or anti-independence positions in the context of a referendum, and thus allows us to identify pro- and counter-attitudinal comments with high accuracy.

The 735 Twitter discussion threads were randomly drawn from a larger sample of 188,839 threads, which were collected from the three most popular hashtags—#9N, #9N2014, and #cassolada—used to support, protest against, or simply report about the Catalan referendum for independence. The threads included the whole discussion tree, from the initial post to all replies, including replies to replies (see Appendix D for details about data collection). To ensure a sufficient representation of different thread types, a random sample was drawn selecting 245 threads of length 0 (original post only), 245 threads with at least one reply where the original poster never replied, and 245 threads where the original poster replied at least once. There were 4,153 total tweets in the sample, contributed by 2,412 unique users.

---

<sup>6</sup> The data were originally collected for a study on protest-related discussion behaviors (Gil-López & Shen, 2021), containing full conversation threads that allow for an in-depth analysis of conversational exchanges on Twitter, beyond retweeting and liking. Including international data allows us to investigate the phenomena outside the United States.



**Figure 5. Structure of Twitter discussion threads.**

Based on this corpus of threads, content analysis was conducted independently by two coders. The content analysis accounted for (1) all user Twitter profiles about their position toward independence (1,300 pro, 516 moderate or undefined, and 596 against), and (2) for the presence or absence of incivility in the main tweet (463 tweets coded as uncivil; 3,690 civil). To assess reliability, two coders jointly coded 10% of the sample (25 conversations of each category); Krippendorff's alpha was .75 for incivility and .82 for user position. Figure 5 shows an example conversation thread with one original post and six replies, two of which are first-order replies.

### **Measures**

#### *Dependent Variables*

*Pro-attitudinal responses* were measured based on the alignment between the original poster's issue position (107 opposing vs. 430 supporting the independence movement) and that of the users who responded to the original tweet (see Appendix E). Responses were coded as pro-attitudinal if the user posting them had the same issue position as the original poster. We also counted pro- versus counter-attitudinal first-order responses, which were replies to the original post. First-order responses are important because they are direct reactions to the original post, rather than to subsequent responses.

### *Independent Variables*

*Original poster's issue position*<sup>7</sup> was coded from users' Twitter profiles and recent posts, which measured a user's position toward Catalan independence as pro or against. This variable was used as a reference to assess the pro- or counter-attitudinal nature of replies relative to the original tweet. Discussion threads initiated by users whose profiles could not be clearly coded as pro or against independence were removed (598 or 81.36% were initiated by users with a clear issue position).<sup>8</sup> *Original post's incivility* was coded as a dichotomous variable indicating whether the first tweet of any given thread was uncivil (44 versus 691 or 94% civil), following Coe et al.'s (2014) operational definition.

### **Analysis**

A fundamental distinction between experimental and observational data about the operationalization of attitudinal congruence was key in the choice of an analytical strategy to parallel that of Study 1 as closely as possible: It was impossible to assess the attitudinal congruence of an original post in a naturalistic setting for those cases where there was never a response. We could characterize only those instances of commenting that did take place. To ensure the robustness of our results, we considered multiple approaches to testing each hypothesis.

### **Results**

H1b predicted that there would be more pro-attitudinal than counter-attitudinal responses to a post. First, the chi-square analysis examined the proportion of pro- versus counter-attitudinal responses. There was a significantly higher proportion of pro- than counter-attitudinal responses, 874 against 339 ( $\chi^2(1, n = 1,213) = 235.96, p < .001$ ), supporting H1b. Such effects were stronger for original posts supporting independence (see Appendix F). Second, to control for possible confounds, such as number of followers and post incivility, we ran a series of negative binomial regressions predicting the total number of first-order replies based on the original poster's issue position (see Table 1). The coefficient for original poster's position was positive (0.468,  $p < .01$ ) when predicting replies from pro-independence users,

---

<sup>7</sup> We also coded for positive or negative evaluations of the movement for independence reflected on the original post and assessed each individual response as pro- or counter-attitudinal. There was a strong correspondence between the individual-response and user levels of measurement ( $r = .60, p < .001$ ). We ran the same analyses with this variable and original poster issue position for comparison. Results were highly similar for both variables, so we proceeded with just the poster issue position to avoid redundancy.

<sup>8</sup> To ensure the robustness of our findings, we prioritized looking at threads initiated by clear issue positions so that subsequent comments could be reliably coded as pro- or counter-attitudinal. Note, as per the coding scheme (Appendix E), that identifying positions as moderate or undefined did not imply uncertainty about the categorization, which was always reliable (see alphas). In highly polarized debates such as the Catalan independence, positions tend to be not only clearly defined but also expressed rather drastically. Ultimately, we did not find significant differences in the number of direct replies between the discussions used in analyses ( $M = 25.79, SD = 55.15$ ) and those dropped at this step ( $M = 29.44, SD = 61.04; Welch t(1193.2) = -1.5686, p = 0.12$ ).

indicating that pro-independence original posters, compared with anti-independence posters, were more likely to receive replies from users also supporting independence. The sign of the coefficient changed for the model, predicting the number of comments from users against independence ( $-1.511$ ,  $p < .001$ ), indicating that pro-independence original posters, compared with anti-independence posters, were less likely to receive replies from anti-independence users. The two models showed that first-order replies are more likely to be pro-attitudinal than counter-attitudinal, again supporting H1b. The third and final test of H1b compared the proportions of pro- versus counter-attitudinal first replies that original posts received. This test is most stringent theoretically and methodologically, because if a post had at least two comments, previous comments' position would influence subsequent respondents' intention to comment just as the original post's position would. Chi-square tests showed that, overall, pro-attitudinal first comments were significantly more common than counter-attitudinal first comments, 257 against 139 ( $\chi^2(1, n = 396) = 35.16$ ,  $p < .001$ ). H1b was thus supported.

**Table 1. Negative Binomial Regression Predicting Number of First-Order Replies (H1b).**

	Replies from pro-independence users			Replies from anti-independence users		
	Estimate	z		Estimate	z	
Intercept	-2.60	-8.12	***	-2.27	-4.93	***
Number of followers (log)	0.33	10.69	***	0.32	6.87	***
Incivility	0.64	2.88	**	0.33	1.00	
Original poster is pro independence	0.47	2.99	**	-1.51	-7.11	***
Observations	396			396		
AIC	1,378.20			800.49		
Theta	1.08			0.54		

Signif. codes: \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ , †  $p < .1$ .

H2b predicted that there would be more comments on civil rather than uncivil posts. A zero-inflated poisson model first tested whether there was a greater number of first-order replies to civil than uncivil posts, controlling for users' number of followers and original poster's position on independence (Table 2). Incivility was positively related to the number of comments a post received ( $0.468$ ,  $p < .001$ ). However, the number of comments could be influenced by the nature of the comments themselves, over and beyond civility. The zero-inflated portion of the model accounted for the likelihood that posts received at least one comment. Results showed that incivility was not significantly associated with the likelihood of posts receiving at least one comment ( $0.359$ ,  $ns$ ). This set of results failed to support H2b.

**Table 2. Zero-Inflated Poisson Regression Predicting Number of First-Order Replies.**

<i>Count model</i>			
	Estimate	z	
Intercept	-2.193	-15.751	***
Number followers (log)	0.409	35.783	***
Orig. poster is pro-independence	-0.582	-9.789	***
Original post is uncivil	0.468	5.517	***
<i>Zero-inflation model</i>			
Intercept	-3.187	-2.145	*
Number followers (log)	0.007	0.059	
Orig. poster is pro-independence	0.986	1.356	
Original post is uncivil	0.359	0.466	
Observations	598		
Log-likelihood	-1,666.048		

Signif. codes: \*\*\*  $p < .001$ , \*\*  $p < .01$ , \*  $p < .05$ .

Finally, H3b predicted that civil posts would get more pro-attitudinal comments than uncivil posts. Chi-square tests showed that the incivility of original posts was not related to greater or smaller proportions of pro-attitudinal first replies ( $\chi^2(1, n = 340) = 0.24, ns$ ) to original posts<sup>9</sup> (see Table 3). Thus, H3b was not supported.

**Table 3. Raw Counts and Proportion of Counter- and Pro-Attitudinal First Replies Given Original Posts' Incivility**

Users' position on independence	Civil	Uncivil
Pro-attitudinal	238 (60.1%)	19 (4.8%)
Counter-attitudinal	126 (31.8%)	13 (3.3%)
Evaluation of independence movement (message)		
Pro-attitudinal	160 (47.1%)	13 (3.8%)
Counter-attitudinal	151 (44.4%)	16 (4.7%)

### Discussion

Using Twitter behavioral trace data during Catalonia independence referendum, Study 2 tested the influence of message incongruence and incivility of an original tweet on subsequent user engagement in naturalistic settings. Consistent with H1b, pro-attitudinal comments in response to an initiating tweet were more prevalent relative to counter-attitudinal ones. H2b about incivility received no support, as Twitter users

<sup>9</sup> Again, this was assessed at the message level (i.e., congruent movement evaluations) in addition to attitudinal congruence at the user level (i.e., original poster and respondent having the same position on independence). Tests show similar results in that they offer no support for H3b.



were just as likely to engage with uncivil original tweets as they were with civil ones. Finally, engagement with civil posts did not differ across levels of attitudinal congruence, which suggested no empirical support for H3b.

### General Discussion

Relying on an online experiment using a mock Facebook page and an observational study using Twitter data, this project extended prior work on factors influencing individual willingness to engage in political discussions on social media. The two studies focused on two salient issues in two countries, with an aim of shedding light on whether message incongruence and incivility of social media posts promote commenting differently in experimental versus naturalistic settings.

The studies offer several timely insights on how people comment on social media content. In an experimental setting, commenting was not influenced by the stance of the original message alone. Although this finding contradicts some evidence that people are more likely to comment on counter-attitudinal messages (e.g., Chung et al., 2015), that evidence is based on self-reported intentions and focuses on commenting on online news rather than on content posted by other social media users, a context that is not only more frequent (inasmuch as more people engage on social media than comment on news) but also generates different commenting behavior than news sites (Rowe, 2015; Ziegele et al., 2014). Moreover, counter-attitudinal posts did not trigger anger or third-person effect. Perhaps that is because gun control is a contentious issue that has been debated for a long time, and therefore, citizens are desensitized and believe others are opinionated and will not be swayed by a single post.

In addition, incivility promoted commenting because it induced anger. It thus seems that the prevalence of online incivility does not legitimize it, and citizens still prefer civil exchanges about public issues. However, incivility, at the same time, discouraged users to speak out—probably by creating an unharmonious and “unsafe” atmosphere for discussion (Hershcovis, Ogunfowora, Reich, & Christie, 2017)—which canceled out the impacts of anger. That is, incivility may turn people off so that they do not comment, unless the anger elicited by incivility is strong enough to outweigh the avoidance.

In the experiment, the effects of incivility depended on message incongruence, in line with some recent work (Druckman et al., 2019; Gervais, 2015, 2017). Specifically, individuals were less likely to comment on a *pro-attitudinal* uncivil post than on a similar civil one; but they were equally likely to respond to *counter-attitudinal* uncivil and civil posts. A possible explanation is that pro-attitudinal incivility may generate high levels of sadness (Druckman et al., 2019), which is related to inaction or withdrawal from others (Lazarus, 1991; Roseman, Wiest, & Swartz, 1994), depressing opinion expressions. On the other hand, individuals may already have negative attitudes toward those who have opposing attitudes, with whom they do not want to interact (Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019), regardless of being civil or uncivil.

Study 2 showed that in a naturalistic setting, discussions on social media mainly occurred among like-minded users and that people were more likely to comment on pro-attitudinal than on counter-attitudinal tweets. Individual choice, network homophily, and algorithmic recommendations, as well as

hashtag choice, could all contribute to such congenial debates. Pro-attitudinal comments would then become much more likely than counter-attitudinal ones, explaining Study 2's findings and their divergence from Study 1. This finding may also be because of the unique context of the collected tweets and the close association of the issue of the Catalan independence with longstanding social movements, which, by their nature, encourage like-minded connections and thrive on homophilic interactions (Freelon, 2015). Had we collected tweets around less salient issues that are not directly related to existing social movements, the results may have differed, and we could have found greater rates of commenting on counter-attitudinal posts. Yet, it is also plausible that users interact with similar opinions simply because such interactions are likely to entail lower risks to their self-presentation and the maintenance of their online interpersonal relationships.

Some limitations need to be noted when interpreting the results. For Study 1, because anger was measured after engagement with the post, we could not rule out the possibility of another causal pathway that incivility promoted commenting, which then elicited greater anger. Given the challenges in measuring anger when participants were reading the post and, at the same time, making the Facebook experience as real as possible, this approach was deemed most practical. Future work should assess participants' emotions unobtrusively before they engage with the post.

Another major concern of Study 1 is the external validity. Participants were forced to read the post about gun control. In reality, most people are not interested in politics and are unlikely to see posts about politics in the first place (Allen, Howland, Mobius, Rothschild, & Watts, 2020; Wojcieszak et al., 2021) and thus will not have a chance to comment on them, especially when the topic is not salient. Besides, because participants were aware that they were participating in a research study, they might behave differently. On one hand, participants knew it was a fake environment and their behavior had no impacts on the real world (i.e., the original poster and other users would not see their replies and reactions). Because two motivations for interacting with others are to express oneself and get positive feedback (Lee & Ma, 2012; Park, Kee, & Valenzuela, 2009), people may be less incentivized to comment if these goals could not be achieved in a fake environment.

On the other hand, in the real world, many users self-censor to avoid offending others (Marwick & boyd, 2011; Pew Research, 2021a). In the experiment, that the replies would not be seen by their friends and family might encourage participants to comment because they would not face any negative consequences. Also, participants might be more likely to comment as a result of their awareness of being observed. We cannot disentangle the effects of these factors and determine which one drives behavior. But we suspect cautiously that they may possibly cancel each other out. Moreover, to reduce the influence of these factors and to increase external validity, we did our best to make the website as real as possible, resembling an actual Facebook experience. Also, we partially addressed this limitation by conducting Study 2, which used trace data from Twitter where people engaged in conversations via hashtags.

It should be noted that the naturalistic setting did not allow for testing all the predictions advanced in Study 1, and the tests did not match exactly what was done in Study 1. That said, the two studies focused on the same concepts and the same phenomenon, and hypotheses tested in two studies were highly similar, with the minor variations responding to the different settings. The two studies complement each other and

offer a more comprehensive picture of commenting behavior on social media. Meanwhile, there does not seem to be a better naturalistic way that would allow us to replicate exactly what we tested in Study 1.

Furthermore, the two studies are different in terms of country, topic, and platform, which may, to some extent, cause the distinct findings. Additionally, this project focused on two controversial issues. It is possible that for other less contentious issues, users' expression modes may be different, which remains to be tested in future studies.

Despite these limitations, this project has important implications. The differential effects detected in the two studies further underscore the constraints of extrapolating the findings from controlled experiments to naturalistic settings of actual online behaviors, exposures, and engagements. It is interesting to show that, in an experiment, individuals do or do not comment on counter-attitudinal or uncivil content. Our findings, however, show that such evidence from an "ideal controlled world" may not transpire outside therein. In short, it is research that triangulates controlled designs with actual behavioral indicators from citizens' unconstrained interactions that is most likely to offer a full portrayal of human behavior.

### References

- Allen, J., Howland, B., Mobius, M., Rothschild, D., & Watts, D. J. (2020). Evaluating the fake news problem at the scale of the information ecosystem. *Science Advances*, 6(14), eaay3539. doi:10.1126/sciadv.aay3539
- Andersson, L. M., & Pearson, C. M. (1999). Tit for tat? The spiraling effect of incivility in the workplace. *Academy of Management Review*, 24(3), 452–471. doi:10.5465/amr.1999.2202131
- Arpan, L. M., & Nabi, R. L. (2011). Exploring anger in the hostile media process: Effects on news preferences and source evaluation. *Journalism & Mass Communication Quarterly*, 88(1), 5–22. doi:10.1177/107769901108800101
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132. doi:10.1126/science.aaa1160
- Balcells, J., & Padró-Solanet, A. (2020). Crossing lines in the Twitter debate on Catalonia's independence. *The International Journal of Press/Politics*, 25(1), 28–52. doi:10.1177/1940161219858687
- Binder, A. R., Dalrymple, K. E., Brossard, D., & Scheufele, D. A. (2009). The soul of a polarized democracy: Testing theoretical linkages between talk and attitude extremity during the 2004 presidential election. *Communication Research*, 36(3), 315–340. doi:10.1177/0093650209333023

- Bruns, A., & Burgess, J. E. (2011, August). *The use of Twitter hashtags in the formation of ad hoc publics*. Paper presented at the 6th European Consortium for Political Research General Conference, Reykjavik, Iceland.
- Chung, M., Munno, G. J., & Moritz, B. (2015). Triggering participation: Exploring the effects of third-person and hostile media perceptions on online participation. *Computers in Human Behavior, 53*, 452–461. doi:10.1016/j.chb.2015.06.037
- 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
- David, P., Liu, K., & Myser, M. (2004). Methodological artifact or persistent bias? Testing the robustness of the third-person and reverse third-person effects for alcohol messages. *Communication Research, 31*(2), 206–233. doi:10.1177/0093650203261513
- Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly, 47*(1), 1–15. doi:10.1086/268763
- Druckman, J. N., Gubitz, S., Lloyd, A. M., & Levendusky, M. S. (2019). How incivility on partisan media (de) polarizes the electorate. *The Journal of Politics, 81*(1), 291–295. doi:10.1086/699912
- Eveland, W. (2004). The effect of political discussion in producing informed citizens: The roles of information, motivation, and elaboration. *Political Communication, 21*(2), 177–193. doi:10.1080/10584600490443877
- Evolvi, G. (2019). #Islamexit: Inter-group antagonism on Twitter. *Information, Communication & Society, 22*(3), 386–401. doi:10.1080/1369118X.2017.1388427
- Feldman, L., Stroud, N. J., Bimber, B., & Wojcieszak, M. (2013). Assessing selective exposure in experiments: The implications of different methodological choices. *Communication Methods and Measures, 7*(3–4), 172–194. doi:10.1080/19312458.2013.813923
- Freelon, D. (2015). Discourse architecture, ideology, and democratic norms in online political discussion. *New Media & Society, 17*(5), 772–791. doi:10.1177/1461444813513259
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication, 14*(2), 265–285. doi:10.1111/j.1083-6101.2009.01440.x
- Gearhart, S., & Zhang, W. (2015). "Was it something I said?" "No, it was something you posted!" A study of the spiral of silence theory in social media contexts. *Cyberpsychology, Behavior, and Social Networking, 18*(4), 208–213. doi:10.1089/cyber.2014.0443

- Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, 12(2), 167–185. doi:10.1080/19331681.2014.997416
- Gervais, B. T. (2017). More than mimicry? The role of anger in uncivil reactions to elite political incivility. *International Journal of Public Opinion Research*, 29(3), 384–405. doi:10.1093/ijpor/edw010
- Gil-López, T., & Shen, C. (2021). "Go, vote, and Tweet it": Interactivity in online protest-related discussions about the 2014 Catalan referendum for independence. *International Journal of Communication*, 15, 3864–3886.
- GU Politics. (2019, April 24). *Overwhelming number of Americans frustrated by incivility in politics, but conflicted on desire for "compromise and common ground."* Retrieved from <https://politics.georgetown.edu/2019/04/24/new-survey-overwhelming-number-of-americans-frustrated-by-incivility-in-politics-but-conflicted-on-desire-for-compromise-and-common-ground/>
- Gunther, A. C. (1995). Overrating the X-rating: The third-person perception and support for censorship of pornography. *Journal of Communication*, 45(1), 27–38. doi:10.1111/j.1460-2466.1995.tb00712.x
- Gunther, A. C., & Mundy, P. (1993). Biased optimism and the third-person effect. *Journalism Quarterly*, 70(1), 58–67. doi:10.1177/107769909307000107
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Publications.
- Hershcovis, M. S., Ogunfowora, B., Reich, T. C., & Christie, A. M. (2017). Targeted workplace incivility: The roles of belongingness, embarrassment, and power. *Journal of Organizational Behavior*, 38(7), 1057–1075. doi:10.1002/job.2183
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22(1), 129–146. doi:10.1146/annurev-polisci-051117-073034
- John, N. A., & Dvir-Gvirsman, S. (2015). "I don't like you any more": Facebook unfriending by Israelis during the Israel–Gaza conflict of 2014. *Journal of Communication*, 65(6), 953–974. doi:10.1111/jcom.12188
- Lazarus, R. S. (1991). *Emotion and adaptation*. New York, NY: Oxford University Press.
- Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, 28(2), 331–339. doi:10.1016/j.chb.2011.10.002

- Marwick, A. E., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), 114–133. doi:10.1177/1461444810365313
- Mendelberg, T., & Oleske, J. (2000). Race and public deliberation. *Political Communication*, 17(2), 169–191. doi:10.1080/105846000198468
- Montanaro, D. (2018, November 1). *Poll: Nearly 4 in 5 voters concerned incivility will lead to violence*. Retrieved from <https://www.npr.org/2018/11/01/662730647/poll-nearly-4-in-5-voters-concerned-incivility-will-lead-to-violence>
- Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, 11, 3182–3202.
- Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton, NJ: 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. doi:10.1017/S0003055405051452
- Nabi, R. (2002). Anger, fear, uncertainty, and attitudes: A test of the cognitive-functional model. *Communication Monographs*, 69(3), 204–216. doi:10.1080/03637750216541
- Ohbuchi, K. I., Tamura, T., Quigley, B. M., Tedeschi, J. T., Madi, N., Bond, M. H., & Mummendey, A. (2004). Anger, blame, and dimensions of perceived norm violations: Culture, gender, and relationships. *Journal of Applied Social Psychology*, 34(8), 1587–1603. doi:10.1111/j.1559-1816.2004.tb02788.x
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259–283. doi:10.1177/1461444804041444
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. New York, NY: Penguin.
- Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *Cyberpsychology & Behavior*, 12(6), 729–733. doi:10.1089/cpb.2009.0003
- Pew Research. (2018, July 11). *Activism in the social media age*. Retrieved from <https://www.pewresearch.org/internet/2018/07/11/activism-in-the-social-media-age/>

- Pew Research. (2021a, May 4). *70% of U.S. social media users never or rarely post or share about political, social issues*. Retrieved from <https://www.pewresearch.org/fact-tank/2021/05/04/70-of-u-s-social-media-users-never-or-rarely-post-or-share-about-political-social-issues/>
- Pew Research. (2021b, September 20). *News consumption across social media in 2021*. Retrieved from <https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/>
- Phillips, T., & Smith, P. (2004). Emotional and behavioural responses to everyday incivility: Challenging the fear/avoidance paradigm. *Journal of Sociology, 40*(4), 378–399. doi:10.1177/1440783304048382
- Porath, C., MacInnis, D., & Folkes, V. (2010). Witnessing incivility among employees: Effects on consumer anger and negative inferences about companies. *Journal of Consumer Research, 37*(2), 292–303. doi:10.1086/651565
- Price, V., Cappella, J. N., & Nir, L. (2002). Does disagreement contribute to more deliberative opinion? *Political Communication, 19*(1), 95–112. doi:10.1080/105846002317246506
- Rathnayake, C., & Suthers, D. D. (2018). Twitter issue response hashtags as affordances for momentary connectedness. *Social Media+ Society, 4*(3), 597–610. doi:10.1177/2056305118784780
- Rojas, H. (2010). "Corrective" actions in the public sphere: How perceptions of media and media effects shape political behaviors. *International Journal of Public Opinion Research, 22*(3), 343–363. doi:10.1093/ijpor/edq018
- Roseman, I. J., Wiest, C., & Swartz, T. S. (1994). Phenomenology, behaviors, and goals differentiate discrete emotions. *Journal of Personality and Social Psychology, 67*(2), 206–221. doi:10.1037/0022-3514.67.2.206
- Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research, 49*(3), 399–425. doi:10.1177/0093650220921314
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society, 18*(2), 121–138. doi:10.1080/1369118X.2014.940365
- Stroud, N. J., Feldman, L., Wojcieszak, M., & Bimber, B. (2019). The consequences of forced versus selected political media exposure. *Human Communication Research, 45*(1), 27–51. doi:10.1093/hcr/hqy012
- Su, L. Y.-F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D. (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society, 20*(10), 3678–3699. doi: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. doi:10.1177/2158244020919447
- Vaccari, C., Valeriani, A., Barberá, P., Bonneau, R., Jost, J. T., Nagler, J., & Tucker, J. A. (2015). Political expression and action on social media: Exploring the relationship between lower-and higher-threshold political activities among Twitter users in Italy. *Journal of Computer-Mediated Communication, 20*(2), 221–239. doi:10.1111/jcc4.12108
- Williams, H. T., McMurray, J. R., Kurz, T., & Lambert, F. H. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change, 32*, 126–138. doi:10.1016/j.gloenvcha.2015.03.006
- Wojcieszak, M., de Leeuw, S., Menchen-Trevino, E., Lee, S., Huang-Isherwood, K. M., & Weeks, B. (2021). No polarization from partisan news: Over-time evidence from trace data. *The International Journal of Press/Politics, 19*401612211047194. doi:10.1177/19401612211047194
- Yang, J., Barnidge, M., & Rojas, H. (2017). The politics of “Unfriending”: User filtration in response to political disagreement on social media. *Computers in Human Behavior, 70*, 22–29. doi:10.1016/j.chb.2016.12.079
- Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication, 38*(1–2), 98–139. doi:10.1080/10584609.2020.1785067
- Ziegele, M., Breiner, T., & Quiring, O. (2014). What creates interactivity in online news discussions? An exploratory analysis of discussion factors in user comments on news items. *Journal of Communication, 64*(6), 1111–1138. doi:10.1111/jcom.12123
- Ziegele, M., Weber, M., Quiring, O., & Breiner, T. (2018). The dynamics of online news discussions: Effects of news articles and reader comments on users’ involvement, willingness to participate, and the civility of their contributions. *Information, Communication & Society, 21*(10), 1419–1435. doi:10.1080/1369118X.2017.1324505