



Article

# Civilizing social media: The effect of geolocation on the incivility of news comments

new media & society

1–21

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/14614448231218989

journals.sagepub.com/home/nms



**Yufan Guo**

Tsinghua University, China

**Yuhan Li**

University of Michigan, USA

**Tian Yang** 

The Chinese University of Hong Kong, Hong Kong SAR

## Abstract

Many social media affordances can affect the quality of online discourse, but such an effect remains understudied for the visibility of geolocation, which is available on most social media platforms. We looked at the event in which Weibo started to display users' IP locations on 28 April 2022, which was supposed to reduce incivility as the deindividuation hypothesis predicted. Leveraging a natural experiment, we examined the effect of IP location visibility, with special attention to COVID-19-related news posts and location-based, uncivil name-calling. We found that displaying the IP location in the comments section increased location-based incivility, as geolocation can function as an effective cue that signals ideological affiliation and fuels conflicts between users holding different political positions on the Chinese Internet. Meanwhile, we characterized a moderating effect of audience size on this decivilizing effect. Our study suggests that diverse social media affordances can fuel group identification and facilitate intergroup behaviors.

---

## Corresponding author:

Tian Yang, School of Journalism and Communication, The Chinese University of Hong Kong, Humanities Building, Shatin, Hong Kong SAR.

Email: tyang@cuhk.edu.hk

## Keywords

Affordance, deindividuation, incivility, location, social identity, social media

Incivility poses threats to political discussion and online discourse, which are crucial to the health of political communication and citizenry (Anderson et al., 2014; Gastil, 2008; Habermas, 1987; Papacharissi, 2004). Thus, scholars have long studied how social media affordances can shape online discourse (Esau et al., 2017; Neubaum and Weeks, 2023) and, particularly, reduce online incivility (e.g. Jaidka et al., 2019). As an affordance ubiquitous on most social media platforms (e.g. Twitter, Facebook, and Instagram), geolocation may also affect public discussion (Lane, 2019), but its impact on conversational styles, especially the use of toxic languages, remains largely underexplored.

In this study, we assessed one particular affordance of geolocation: IP location. Affordances refer to object's potentials for action (Gibson, 1977). As a relational concept linking material features and human agency, affordances can enable or constrain some uses of technology (Evans et al., 2017), which may affect users' attitudes and behaviors (e.g. Jaidka et al., 2019; Wang and Sundar, 2022; Zhou, 2021).

We argue that geolocation influences online incivility, as it can function as an effective cue that signals various identities, including ideological affiliation, which facilitates group identification and intergroup animosity. This mechanism is not unique to geolocation. For example, Settle (2018) found that words, videos, and images of lifestyles posted on social media, which seem irrelevant to ideological positions and political attitudes, can serve as partisan identifiers to fuel polarization (Settle, 2018).

We examined the causal effect of geolocation affordance on one type of online incivility. Generally, incivility is the expression of disrespect and offensiveness toward others (Coe et al., 2014; Kim et al., 2021). Given the relevance to geographical location, we particularly focused on a special form of uncivil speech, *location-based name-calling*, which is a direct result of the presence of geographical identity on social media.

Furthermore, we tested the contingencies of such an effect and focused on one critical factor, audience size, which can shape online discourse, as shown in many previous studies (Gil-Lopez et al., 2018; Kim, 2020). Although prior studies have demonstrated that larger audiences are positively associated with more online expressions (Barnidge et al., 2016; Eveland and Hively, 2009), we know little about how the perceived audience size affects people's online incivility and how this effect moderates the impact of geolocation.

We employed a natural experiment to analyze the event in which Weibo, a major Chinese social media platform, started to display all users' IP locations on 28 April 2022. We focused on a consecutive 25-day period and analyzed more than 160,000 news comments on COVID-19 news posts, which revealed that displaying the IP location in the comments section led to more uncivil location-based name-calling. Our study provides an empirical assessment to show how an expanded array of social media affordances functions as identity cues in a hyper-political environment, which casts light on digital governance and content moderation.

## Literature review

### *Geolocation on digital platforms*

As a commonly used affordance, users' geolocation has been incorporated by most major social media platforms for various purposes, such as geotagging, check-in, and self-presentation. Users make use of their geographical information for dating nearby others (Blackwell et al., 2015), cultivating social capital (Hjorth and Gu, 2012), organizing collective actions (Su et al., 2022), and performing their identities (Schwartz and Halegoua, 2015). Previous studies have mostly paid attention to volunteered geographical information disclosed by users themselves yet largely overlooked another form of location presentation, which is automatically captured and exhibited via geo-sensors of mobile devices (Hoffmann and Heft, 2020).

Despite a large body of research on geolocation, limited studies have shed light on the role of geolocation in online discussions. One exception is Peng's (2021) comparison of comments sections on two news websites. Peng found that, in the comments of a news article covering a scandal in a Chinese province, regional discrimination was more prominent on the website that displayed commenters' IP locations. However, we argue that location-based incivility is not limited to the discussion of negative local news. Rather, it may go beyond news content with geographic relevance. In this study, we investigated whether and how the presence of geolocation as a social identity shapes news discussion on social media.

### *IP location and antisocial deindividuation*

We highlight two lines of theories to consider the possible impact of displaying IP location on online incivility. First, *antisocial deindividuation* posits that anonymity leads to deindividuation and may induce anti-normative behavior (Mann et al., 1982; Perfumi et al., 2019). Deindividuation refers to a situation in which individuals do not see themselves as individuals and decrease self-evaluation (Festinger et al., 1952). In the online environment, where people are visually anonymous, do not necessarily reveal their personal identities, and experience deindividuation (Sproull and Kiesler, 1986), individuals can exercise their autonomy without much fear or shame of social disapproval and conduct more antisocial communicative behaviors. The concept of *toxic online disinhibition* articulates a similar process: A highly indistinguishable person can avoid responsibility for his or her negative actions because of the elusive linkage between personal identity and online behaviors (Suler, 2004).

Several studies have provided empirical evidence in various contexts that a higher level of anonymity is associated with higher incivility in the comments section of political news (Knustad and Johansson, 2021; Rowe, 2015; Santana, 2014). For example, Rowe (2015) found that a media website had more uncivil comments than its Facebook page, on which the personal identities of commenters were revealed. In contrast, a decrease in anonymity, which exerts more normative social pressures on online users, will lead to a chilling effect (Penney, 2016). They should hesitate to post uncivil words due to the fear of being surveilled by others and the power that may lead to being

personally attacked. According to these studies, the incivility on social media will decrease, as displaying IP locations induces a lower level of personal anonymity.

### *IP location as an ideological identity marker*

However, *social identity theory* (SIT; Tajfel, 1978) suggests an alternative direction of the effect. In a computer-mediated communication environment, one extension of SIT, namely, the *social identity model of deindividuation effects* (SIDE), may help to further elucidate intergroup behaviors. The theory illuminates how behavior criteria can shift from the individual level to the social level of identification when personal identity information is limited (Reicher et al., 1995). Similar to the deindividuation theory, one precondition of the SIDE is visual anonymity, or people being unable to see or be seen by others due to the lack of visual cues (Scott and Rains, 2020). The SIDE proposes that visual anonymity can make the characteristics of the group as a whole more salient (Postmes et al., 2001). However, different from the classical deindividuation theory, the SIDE predicts that people will conform to specific group norms of salient social identities signaled by those online cues and behave accordingly (Postmes et al., 2001; Reicher et al., 1995) instead of exhibiting an idiosyncratic personal identity. A recent study used the SIDE to explain the contagion of national identity narratives on social media (Chen et al., 2023): Among individually anonymous users, the use of national identity language predicted the approval of ingroup stance and imitation of ingroup language in everyday social media discussions.

Some studies on the SIDE further distinguished personal anonymity and social anonymity (Jaidka et al., 2022; Reicher et al., 1995: 177). Personal anonymity refers to the degree to which one's personal identity is undetectable, whereas social anonymity characterizes the degree to which one's social category can be identified (Reicher et al., 1995). A recent study found that these two together exerted an interaction effect (Jaidka et al., 2022), which suggests that the effect of social identifiability is conditional on other contextual factors.

Particularly, while the SIDE suggested an identification process, the specific group behaviors are contingent upon certain intergroup contexts: Interactions between opposed groups were often more negative than those between ingroup members (Marchal, 2022). Meanwhile, the faction of ingroup members also matters—a large group size reduces the need to emphasize and secure one's identity, which decreases the motivation to post uncivil comments (Rains et al., 2017).

On social media, various affordances can function as identity cues to facilitate such a social identification process. In addition, even in a visually anonymous environment, people draw inferences about the possible social identities of other users from their online traces, including, but not limited to, profile pictures and pages (Munger, 2017), the content they post (Settle, 2018), and network associations (Fox and McEwan, 2017). Once people recognize the social identity of other individuals, they react accordingly. According to SIT, these group-based behaviors generally include ingroup favoritism and outgroup derogation: Individuals prefer to benefit and cooperate with ingroup members (Balliet et al., 2014) while making negative evaluations about the outgroup and engaging

in hostile behaviors under certain conditions (Marchal, 2022). The process further augments identification and shapes intergroup dynamics.

In our case, the presence of an IP location should indicate one's regional identity. In addition, like other affordances (e.g. Settle, 2018), it sometimes goes beyond revealing one's physical location to indicating one's ideological affiliation. Sorting among various identities further highlights group differences and exacerbates intergroup tensions. This is especially the case in China: Similar to many Western democracies, the ideological spectrum of online users in China is far from constrained but dispersed and along multiple dimensions (Pan and Xu, 2018). Besides a sizable group promoting liberal ideologies online (Yang, 2009), with the prominence of right-wing populism and xenophobia on Chinese social media (Yang and Fang, 2023; Zhang, 2020), ingroup solidarity and outgroup exclusion motivated by cyber-populism have risen (Guan et al., 2021). The conspiracy narratives of so-called "foreign forces"<sup>1</sup> in state propaganda have also exacerbated tensions between groups holding various ideologies in Chinese cyberspace. Digital platforms afford a variety of cues and channels for expressing political disagreements (Han, 2018a), and news comments on social media provide an arena for users to interact with groups that hold different positions. In such a context, whether one is abroad or not has been a strong cue to signal one's ideological identity. For many Chinese social media users, accounts with overseas locations from "unfriendly" areas are likely to hold very different viewpoints.

Moreover, there is a close connection between ingroup identity and intergroup conflict (Brewer, 2001). Especially regarding groups with very different ideological affiliations, the salient group identity can enable anonymous users to behave increasingly impolitely or aggressively against people they assume come from the ideologically opposing camp. Aggressive intergroup behaviors may further trigger a "spiral of incivility" to promulgate uncivil feedback from the other side (Chen and Lu, 2017; Kim et al., 2021).

Given the sorting between geolocation and ideological identities, the presence of IP location may fuel group identifications and intergroup conflicts (domestic vs overseas) in news comments on social media, especially regarding public issues about which people hold a variety of opinions and no single group clearly dominates the conversation.

In short, considering the competing predictions of *antisocial deindividuation* and the SIDE regarding the effects of IP location, we propose the first research question:

*RQ1.* How does the presence of users' IP locations affect location-based name-calling in news comments?

### *Role of audience size in incivility*

As an important contextual factor, audience size matters to the quality of discussions and people's interactive patterns (Choi et al., 2018). Audience size is about the number of people with whom someone communicates (either real or imagined), which can influence what people discuss and how they interact (Barasch and Berger, 2014). However, prior studies on the effect of audience size on incivility are mixed. First, some

scholars have suggested that those who have a larger audience tend to be more civil in discussions. In a larger political discussion network, people have a higher number of weak ties (Gil De Zúñiga and Valenzuela, 2011), which should provide access to a diverse pool of perspectives and political viewpoints. The chance of hearing various perspectives nurtures a high level of political tolerance (Mutz, 2002). Meanwhile, people with a larger network exhibit more civil behaviors in online discussions with higher visibility (Kim, 2020): If users perceive that they are exposed to larger audiences and are more likely to hear different opinions in discussion networks, they tend to self-regulate their political expressions to avoid conflicts with unintended audiences (Klofstad et al., 2013; Vraga et al., 2015).

These aforementioned studies, however, did not consider the ephemeral nature of many online networks, where higher levels of incivility may be observed when people have opportunities to interact with a wider audience. When facing a larger audience, it is easier for people to encounter more unfamiliar people. That is, people can easily disconnect from weak ties because of the lower psychological cost (Yang et al., 2017), especially when users encounter political disagreements in online discussions (John and Dvir-Gvirsman, 2015). People can be disrespectful and then cut off ties when they take offense from unfamiliar people in online discussions (John and Dvir-Gvirsman, 2015). As a result, those in large discussion networks are not afraid to freely express their opinions because they are not familiar with most of their audience, and it costs little to maintain these weak ties. Thus, a larger audience can be associated with more intolerance, as well as more uncivil and toxic language.

Overall, given the mixed evidence in previous studies, we propose a research question:

*RQ2.* How does the audience size of news posts affect the level of location-based name-calling in news comments?

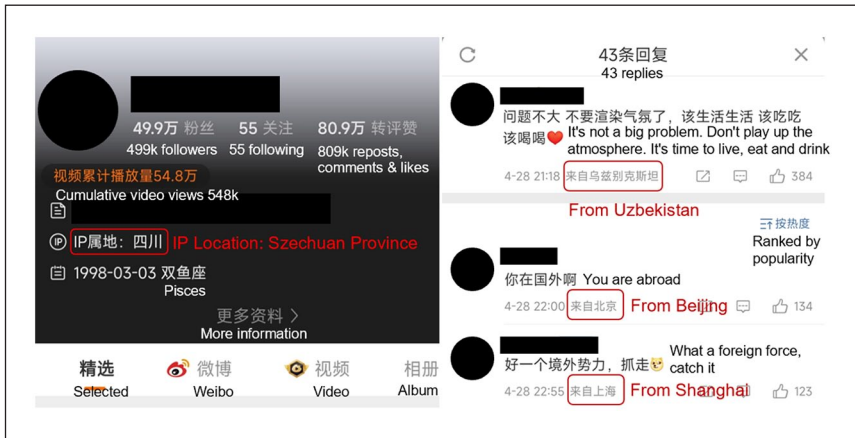
As audience size may shape the interaction settings and configure the dynamics of deindividuation and group identification—two critical mechanisms by which IP location affects incivility—we expect a moderating effect of audience size here. Given the limited evidence we have on this relationship, we propose a research question on how audience size conditions the effect of IP location on incivility.

*RQ3.* How does the audience size of media outlets moderate the relationship between displaying users' IP locations and the level of location-based name-calling in news comments on social media?

## Methods

### *Natural experiment setting*

We chose Weibo, the Chinese version of Twitter,<sup>2</sup> as the context of this study. As one of the most influential social media platforms in China, Weibo had accumulated 582 million monthly active users by the second quarter of 2022 (SinaTech, 2022), when IP



**Figure 1.** Example of user homepages and comments sections on Weibo.

Note. The red words are the displayed IP locations. We also included English translations of other information.

locations started to be displayed. In the last decade, Chinese Internet companies have constantly struggled to find a balance between the state's information control and commercial interest (Han, 2018a; Jia and Han, 2020). The state's tightening regulation on online speech promoted a "civilizing process" to regulate so-called "uncivil" online discussion (Yang, 2018), which essentially strengthened social control. Internet service and content providers had to deploy a series of monitoring and censoring mechanisms to meet the expectations of the authorities. Social media platforms are also dedicated to mitigating the negative impact of censorship on user experiences and commercial profits. Hence, platforms frequently instituted changes in their technical designs to respond to the state's regulations. For example, Weibo introduced a real-name registration policy<sup>3</sup> and the user credit system<sup>4</sup> to promote the identifiability and surveillance of online content to comply with government regulations (Jia and Han, 2020).

We used a natural experiment to study the impact of disclosing geolocation. On 28 April 2022, Weibo started to reveal all users' IP locations, both on the users' homepages and in the comments sections (see Figure 1). Domestic locations were displayed at the provincial level, whereas foreign locations were located at the country level. Weibo (2022) claimed that the adjustment aimed to "maintain a healthy and orderly discussion atmosphere" and "ensure the authenticity and transparency of the communicating content," which especially aimed to prevent the dissemination of misinformation. We treat displaying IP location as an exogenous intervention in an online discussion, which provides a chance to examine its effect on incivility, with a special focus on location-based uncivil name-calling.

## Data

The process of data collection is displayed in Figure 2. Referring to Weibo's official list (<http://v6.bang.weibo.com/xmt>), which covers the most popular accounts in the category



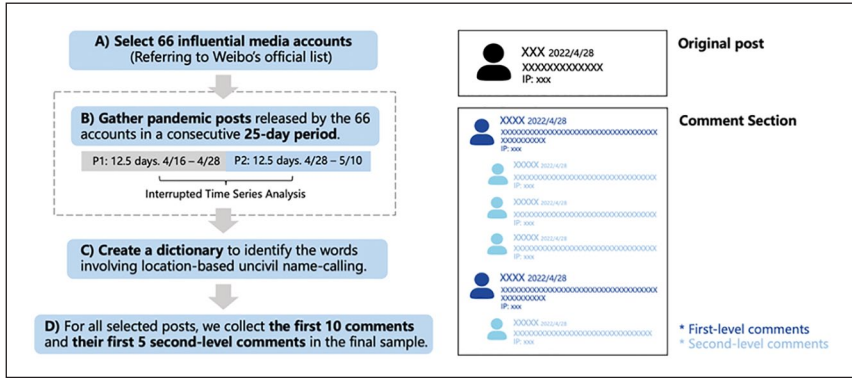


Figure 2. Process of data collection.

of media, we first compiled a list of the 66 most influential media accounts, which consist of official media (newspapers, magazines, and media websites), commercial media, and self-media<sup>5</sup> (see Table A1 in Supplementary Information). This list includes most major news sources on Weibo.

We gathered data over a consecutive 25-day period: 12.5 days before the event as the pre-intervention period (Period 1: 16–28 April) and 12.5 days after the event as the post-intervention period (Period 2: 28 April–10 May). Comments on posts published by the selected accounts during the pre-intervention period were treated as our control group, and comments produced in the post-intervention period became our treatment group.

The COVID-19 pandemic and related social policies were chosen as the news topics we looked at in this study. Given that the Chinese government has implemented radical and controversial measures to prevent the evolving pandemic since 2020, heated debates about the effects, necessity, and legitimacy of pandemic policies have been prominent. Given the stark differences in public policies on pandemic measures between China and most other countries, it was believed that people overseas held opinions that differed from those of the domestic population (e.g. Qian et al., 2021). Furthermore, the COVID-19 pandemic was an ongoing and generally stable news topic during the time window, and pandemic policies in China were stable over the period.<sup>6</sup> The assignment of the treatment on displaying IP location can be treated “as-if” random for posts and comments related to this issue. See Supplemental Appendix IV for additional information about the COVID-19 topic and its context.

After collecting all posts from these accounts during these two periods, we updated the Weibo keyword list developed by Shen et al. (2020) based on our observations (see Table A2 in Supplementary Information) to automatically identify 17,285 posts discussing COVID-19-related issues.

Weibo comments were displayed in a two-layered structure (see Figure 1 and Figure 2). We included the first 10 comments on each post (first-level comments) and their first 5 responses (second-level comments) in the final sample. Therefore, each post would have a maximum of 60 comments to be analyzed. The ranking of comments in both layers was largely determined by the likes received, which means that the comments we sampled



were the most popular and influential ones. They were viewed by and engaged with the most people. All comments released after May 10 were discarded. Overall, we obtained 162,887 comments.<sup>7</sup>

## Measurement

**Online incivility.** This study focused on a special type of incivility: location-based, uncivil name-calling. As Lee et al. (2019) argued, it is difficult to develop a complete scheme encompassing all forms of uncivil expression using a computational approach. More importantly, incivility in online political discussions is rooted in a specific platform culture, so the strategy of operationalization should be highly contextualized and constantly updated. Based on Mandarin Chinese profanity in Wikipedia<sup>8</sup> and researchers' long-term digital observations of Weibo discourse, we created a dictionary to identify the words involving location-based uncivil name-calling (see Table A3 in Supplementary Information).

The 96-word dictionary mainly comprises three kinds of uncivil terms. The first category involves the mutual stigmatization of countries, regions, and people between domestic and overseas users. One example is "guizi" (鬼子), which means "the devil" in English. It refers to Japanese invaders in World War II and has become a pejorative term targeting Japanese people in online discussions. Another example is "zhiqu" (支蛆), an insulting and racist term targeting Chinese people, especially those who support the current Chinese government. The second category includes negative words related to political disagreement. The term "wushiwan" (五十万), for instance, literally translates to "half a million dollars." It is a satirical expression used to label users who support the US position and interests by accusing them of receiving illegal funding from the US government. The third part also denotes overseas entities but is used for political satire aimed at domestic users. An example of this type is the term "meidie" (美爹), literally meaning "American daddy," which is used to shame pro-US users. Comments containing any word in the dictionary were regarded as "uncivil comments" and coded as 1 (0.86%), whereas others were coded as 0 (99.14%).

For validation, we randomly selected 200 each of civil and uncivil comments from our dataset and invited two graduate students familiar with online political discussions on Weibo to code these 400 comments. We obtained acceptable reliability (Krippendorff's  $\alpha=0.9$ ), which supported the validity of our dictionary in gauging uncivil speech. See Supplemental Appendix II for more details.

**Treatment.** We coded a binary variable indicating whether the IP location was displayed for the comment (0=pretreatment, 53.51%; 1=posttreatment, 46.49%).

**Audience size.** We created an account-level variable, *audience size*, which was measured by the number of followers of each account. We log-transformed and further standardized the values ( $M=16.39$ ,  $SD=1.70$ ).

**Time.** As an important variable used in our model (see below), time was coded as a continuous variable indicating the temporal distance from the beginning of our observation.

We looked at the exact second of the commenting time and transformed the relative temporal distance into a value ranging from 0.00 (16 April, 00:00) to 25.00 (10 May, 24:00).

### Analysis strategy<sup>9</sup>

We conducted an interrupted time series (ITS) analysis to determine whether displaying a user's IP location can affect the incivility of news comments on social media. ITS analysis is a variation of regression discontinuity design, where the running variable is time. This method allows researchers to make causal inferences when randomized trials are impractical (Morgan and Winship, 2015). It requires data collected at multiple instances before and after the introduction of an intervention. In our study, the intervention was Weibo's policy of displaying IP locations starting on 28 April 2022. Thus, ITS analysis can test whether the outcome observed in the post-intervention period is significantly different from that observed before the intervention, which allows us to characterize the effect of the intervention (Ramsay et al., 2003). Because our outcome variable was binary, we used generalized linear mixed-effects models to test our hypotheses.<sup>10</sup> The following equation describes the model used to characterize the main effect ( $j$  means the media account to which each comment belongs)

$$\Pr(\text{Incivility}_{ij}) = \text{logit}^{-1}(\beta_{0j} + \beta_1 T + \beta_2 \text{Treatment}_t + e_{ij})$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

In the equation,  $\text{Incivility}_{ij}$  refers to the incivility (1 or 0) of the comment posted at time  $t$  on media outlet  $j$ , whereas  $T$  refers to time, and  $\text{Treatment}_t$  denotes a dummy variable indicating the presence of IP location.

## Results

RQ1 asks whether displaying IP location would decrease or increase the incivility of new comments. As shown in Table 1, the treatment in Model 1 significantly predicted an increase in incivility ( $b=0.289$ ,  $se=0.104$ ,  $p=.005$ ). Compared with the pretreatment period, the odds of location-based name-calling increased by 33.5% after the IP location was displayed on Weibo.<sup>11</sup>

To better present the decivilizing effect of displaying IP locations, we visualized the predicted possibilities of location-based name-calling (Figure 3). Before Weibo's new policy on IP location, the predicted possibility of uncivil words showed a slight downward trend. However, it sharply increased from 0.35% to 0.46% when the intervention occurred on 28 April, an increase of more than 30% compared with the pretreatment period.

Regarding the effect of audience size, RQ2 asks the relationship between the audience size and the level of news comment incivility. To answer the question, we added the term audience size and its interaction with the treatment into the model. The result in Model 2 suggests that a larger audience significantly predicted a lower level of incivility ( $b=-0.922$ ,  $se=0.147$ ,  $p<.001$ ). A one standard deviation increase in the logged size of

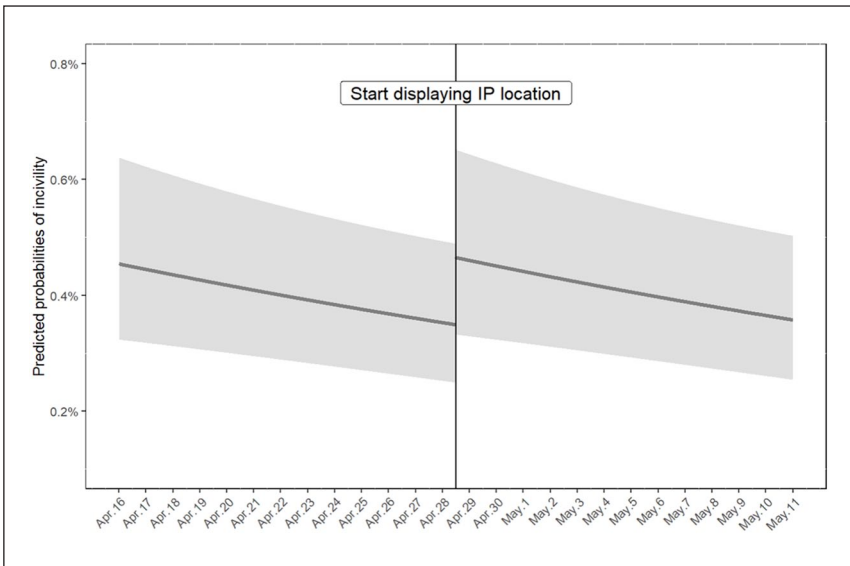
**Table 1.** Generalized linear mixed-effects models predicting incivility.

|                             | Model 1              | Model 2              |
|-----------------------------|----------------------|----------------------|
| Intercept                   | -5.390***<br>(0.174) | -5.449***<br>(0.145) |
| Treatment                   | 0.289**<br>(0.104)   | 0.451***<br>(0.110)  |
| Time                        | -0.021**<br>(0.008)  | -0.020**<br>(0.008)  |
| Audience size               |                      | -0.922***<br>(0.147) |
| Treatment × audience size   |                      | 0.273***<br>(0.061)  |
| $\sigma^2$                  | 3.29                 | 3.29                 |
| $\tau_{00}$ (media account) | 1.35                 | 0.76                 |
| N (level 1)                 | 162,887              | 162,887              |
| N (level 2)                 | 66                   | 66                   |
| AIC                         | 14,765.8             | 14,726.8             |
| BIC                         | 14,805.8             | 14,786.8             |

AIC: Akaike information criterion; BIC: Bayesian information criterion.

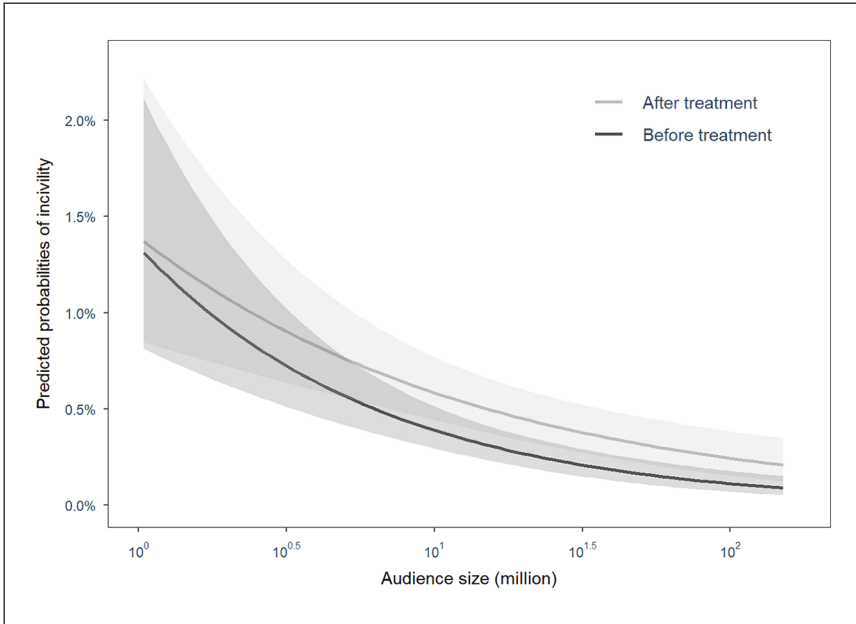
Unstandardized coefficients with standard errors are in parentheses.

\*\* $p < .01$ . \*\*\* $p < .001$ .



**Figure 3.** Treatment effect of displaying IP location on incivility.

Note. The shaded areas represent the 95% confidence intervals of the estimates.



**Figure 4.** Predicted probabilities of incivility by audience size.

Note. The shaded areas represent the 95% confidence intervals of the estimates. The range of the audience size was 1.5 standard deviations above and below the mean.

the potential audience would be translated into a decrease in the odds of location-based name-calling by 60.2%.<sup>12</sup>

RQ3 explores the heterogeneity of the treatment effect moderated by audience size. The results of Model 2 showed that the interaction item between the treatment and audience size significantly affected location-based name-calling ( $b=0.273$ ,  $se=0.061$ ,  $p<.001$ ). We visualized the moderating effect in Figure 4. In general, the gap between the two curves became wider as the audience size grew larger. Hence, we can conclude that a larger audience size predicted an augmented decivilizing effect of displaying the IP location. In other words, when people post comments on accounts with fewer followers, whether the IP location is visible matters less. The effect we identified—displaying IP location facilitates more uncivil comments—is more prominent among news accounts that have more followers.

### *Robustness checks, placebo test, and mechanism test*

We further conducted a series of robustness checks using different model specifications. First, considering the special media system in China, the audience size is likely to be related to media ownership (i.e. state-owned media or others), which may account for the effect we observed. Thus, we controlled for media ownership by adding it into the model (see Table A4 in Supplementary Information). Second, we used fixed-effects models, which excluded time-invariant unobserved differences between media accounts, to test

whether our conclusions were still the same (see Table A5 in Supplementary Information). Third, we considered the nonlinear effect of time trend by adding another item,  $Time^2$ , into the baseline model (see Table A6 in Supplementary Information). Fourth, we estimated the treatment effect by adjusting the time window observed (from 25 to 17 days). The comments collected within 8 days before and after 28 April were regarded as a new dataset for testing our hypotheses (see Table A7 in Supplementary Information). Fifth, we added the interaction item between *Treatment* and *Time* into mixed-effects models (see Table A8 in Supplementary Information). All the results of the above tests were similar to the findings reported above.

In addition, we ran a placebo test to check whether the effect is indeed explained by the policy change on IP location. 22 April, a week before the real intervention, was treated as a pseudo-intervention for conducting such a placebo test. The original pre-intervention period was divided into two groups: a control group (16–21 April) and a “treatment” group (22–27 April). As there was no real intervention at that time, we expected that the effect of “treatment” would disappear when we ran our models again on this dataset. We found that the main effect of displaying IP location became insignificant (see Table A9 in Supplementary Information). However, the effect of the interaction item did not disappear.

We also conducted an additional mechanism test to further support that it is the association between geolocation and ideological positions that plays a key role in facilitating the rise of incivility. The test revealed that “between overseas and China” posttreatment second-level comments showed the highest incivility level, significantly higher than “between different provinces” posttreatment second-level comments. The finding suggests that it is the ideological positions extrapolated from geolocation that mainly explain the increase in incivility. See Supplemental Appendix III for more information.

## Discussion

### *Decivilizing effect of IP location*

By displaying users' IP locations, Weibo endeavored to civilize online discourse. We found that, after the rollout of the IP location policy on Weibo, the location-based incivility of news comments on COVID-19 actually increased (RQ1). This effect contrasted with the antisocial deindividuation hypothesis but aligned with the SIDE. The dynamics of group identification, fueled by the prominence of group identity and personal anonymity in a visually anonymous online environment, facilitated intergroup animosity and particularly, the use of uncivil language. By characterizing outgroup members in terms of an outgroup stereotype (Reicher et al., 1995), individuals deem outgroup members political rivals by simply referring to the IP location. As our case indicates, ideological disagreements around pandemic issues extrapolated from geolocation information have driven people to use location-based, uncivil name-calling. This echoes previous studies on how displaying social identity (e.g. political affiliation) on the Internet can fuel intergroup conflicts and polarization (Marchal, 2022; Rains et al., 2017).

This effect supports the potential for a wide range of social media affordances to function as identity markers, which can fuel group identification and intergroup

behaviors. This is not unique to geolocations. Settle (2018) observed that social media users usually post a variety of content on social media. Not only do all kinds of content have the potential to be politicized, given the sorting between various identities, but almost all social media affordances can be interpreted politically. Besides some explicit opinion expressions and political engagement, other elements, including the use of certain products, could all implicitly signal ideological identities (Settle, 2018). As we see a rise of polarization, politicization, and political sorting in other societies (Lee, 2021; Peterson and Muñoz, 2022), we speculate that many affordances that were not originally designed to reveal political identities, including IP location, can act as *de facto* ideological indicators. They can transform interpersonal interactions into intergroup encounters in online environments.

Our findings also provide practical implications for digital platforms and policymakers by presenting the paradox of affordance adjustment. Internet companies have invested many resources in content moderation (Gillespie, 2018). Compared with manual inspection and automating moderation, technological adjustments, such as displaying IP locations, may be less costly and less sophisticated. However, social media affordances are actually embedded in a broader sociotechnical system whose functions may comprise multiple and complex dynamics (Crawford and Gillespie, 2016). Our study suggests that social media platforms should give more consideration to unintended consequences when they decide to roll out such interventions. A seemingly simple and straightforward approach may have side effects.

### *Audience size as a moderating factor*

Our study also suggests that a larger network is associated with a higher level of civility in user discussions (RQ2). People in larger networks may believe that they have more chances to encounter different views, so they have to take more risks of being judged and surveilled by others. This setting increases their tolerance of opposing opinions (Mutz, 2006) and openness to political disagreements (Moy and Gastil, 2006). These mechanisms may explain why people were less likely to use location-based smear words and name-calling in the comments of posts published by more popular news accounts. This finding indicates that the comments of popular news accounts meet several key ideals of the public sphere (Habermas, 1987): The large network is more conducive to constituting a more inclusive and diverse digital space. These spaces are also likely to be more civil, according to our study. They have the potential for civic engagement and digital governance.

However, the moderating effect of audience size on how IP location affects incivility displays a gloomy image: The decivilizing effect is especially augmented in these large networks (RQ3). Although a big arena makes it possible for people to benefit from diverse opinions, it also increases users' probability of encountering overseas accounts and further fuels intergroup conflicts when discussing politically controversial issues, which undermines the civic potentials that we discussed above.

This moderating effect also casts light on the contingency of the effect of anonymity in computer-mediated communication. In contrast to previous studies (e.g. Halpern and Gibbs, 2013; Oz et al., 2018; Rowe, 2015), our study showed that in a large

network, reducing the level of anonymity increased incivility. This discrepancy may be explained by the fact that the anonymity and the nature of digital networks vary across these contexts. For example, Rowe (2015) compared comments on a large news website and its corresponding Facebook page. The contrast revealed the effect of displaying personal identity, not social identity, which was the focus of this study. The distinction between these two identities has also been documented in several studies (e.g. Jaidka et al., 2022). Furthermore, the social network environment contains abundant weak and personal ties, whereas online websites do not. In summary, the contextual settings of the online environment and the particular form of identifiability may greatly shape the effect of anonymity. The typology and contingencies of the anonymity effect should be further explored.

### *Limitations and future research*

All studies have limitations. Ours is not an exception. First, we only paid attention to the comments of news posts related to COVID-19 issues, which is a controversial topic in our context. Our findings on intergroup conflicts may be conditional on the fact that attitudes toward this topic could be associated with the ideological position often linked to their geolocation. For topics on which the public has largely reached consensus, the sorting between various identities and attitudes may not be prominent, so the effect of IP location may disappear, which should be substantiated in future studies.

Furthermore, we focused only on location-based identity and related incivility, whereas social media affordances can signal other social identities, such as gender, race, and sexuality. The sorting between them and ideologies is prominent in many contexts. For example, in the eyes of populists, women and feminists in China are mostly liberal and favor Western liberal ideologies (Han, 2018b). Scholars can further investigate these identities and how the relevant social media affordances exert similar effects. The findings should help us understand the dynamics of incivility in online discussion.

Moreover, as our research scope was limited to one specific type of uncivil speech, whether this effect is generalizable to other aspects is unknown. Future studies should investigate whether this mechanism has the potential to be applied to other forms of uncivil language and the overall level of incivility.

Meanwhile, this study overlooked various intergroup factors. The decivilizing effect of IP location is contingent on the social sorting of regional and ideological identities (e.g. Qian et al., 2021), as well as the particular intergroup contexts highlighted in previous SIDE studies (Marchal, 2022; Rains et al., 2017). For example, when people perceived a higher fraction of ingroup members in the discussion, they posted more civil content (Rains et al., 2017). This perception is shaped by the technological affordances and general opinion climates on the platform. How these factors together shaped the effect we characterized should be further explored.

Finally, we focused on Weibo, a Chinese social media platform. This particular case shared many similar affordances with other platforms, but whether the conclusion is generalizable needs empirical assessment. As previously discussed, the findings may be contingent on certain contextual factors that can be grasped from a cross-platform comparison. In addition, as a type of news engagement, news commenting behaviors are



always shaped by larger social systems. Whether our findings from China can be applied to other settings with different media systems and state–media relationships should be tested in the future.

## Conclusion

This study investigated the influence of geolocation on online location-based incivility and revealed that displaying IP location on social media increased location-based uncivil name-calling. Moreover, a comment with a larger audience exhibited a lower level of incivility but showed a higher level of the decivilizing effect of geolocation. This study advances our understanding of how social media affordances shape online discourse by functioning as identity markers, fueling group identification, and exacerbating inter-group animosity, which aligns with the SIDE. It sheds light on content moderation, highlighting the need for platforms to address the unintended consequences and to consider the interplay between technological affordances and group dynamics for creating a better online information environment.

## Acknowledgements

We are grateful for feedback from Alvin Zhou, Jinping Wang, Yilang Peng, Hai Liang, and research assistance from Haoshuai Xu, Xinying Tan, and Jiajia Guo.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iD

Tian Yang  <https://orcid.org/0000-0003-2642-4760>

## Supplemental material

Supplemental material for this article is available online. Replication materials can be found at <https://osf.io/4qnte/>.

## Notes

1. “Foreign forces”(境外势力) is a phrase to refer to overseas organizations and people who challenge the dominant power and ideologies in China. In both Chinese Internet and official discourse, many political events and online hotspots are attributed to “foreign hostile forces” (McKune, 2015).
2. Different from Twitter, on Weibo, comments generated by users’ followers will not be displayed on users’ timelines. This means that commenters’ social ties largely cannot affect the dissemination of the commented posts.
3. Under the real-name registration policy, users must provide their real names to the platform.
4. Weibo established the user credit system in 2012 to manage web content and users. Each user is given a starting credit score of 80, which will be increased or decreased based on Weibo’s assessment of the user’s performance. The user’s account will be automatically closed when the credit score falls to 0 (see Jia and Han, 2020).

5. Self-media (also translated as *we media*) refers to noninstitutional content providers on Chinese social media platforms (Fang, 2022), usually distinguished from mainstream media and official media. Here, we define self-media as Weibo accounts operated by individuals or small-scale teams who mainly express their views on current political and social issues. We included a few influential self-media accounts in our analyses, given their prominence in today's Weibo media landscape. See Table A1 for more information.
6. We conducted a Wilcoxon signed-rank test to examine whether the posts related to COVID-19 differed for news accounts we examined before and after the treatment, which were two equal periods. The result revealed no difference ( $V=1144, p=.643$ ). We also tested whether this was the case for a narrow time window (17 days; see robustness section below), which showcased a qualitatively similar result ( $V=1186.5, p=.142$ ). These findings aligned with the “as-if” random assignment assumption here.
7. Not all news posts chose to completely display their responses. To eliminate the impact of censorship of news posts, we removed all the posts selectively displaying their comments and tested the same models with the new dataset. The results did not change (see Table A10 in Supplementary Information). Except in the mechanism testing, we did not differentiate first-level comments and second-level comments given that users often do not follow a strict “reply” structure—for example, they may post a first-level comment to criticize another first-level comment rather than posting a second-level comment.
8. See: [https://en.wikipedia.org/wiki/Mandarin\\_Chinese\\_profanity](https://en.wikipedia.org/wiki/Mandarin_Chinese_profanity).
9. All replication materials, including scripts and data files, can be accessed by an online repository.
10. Mixed-effects models consider differences between media accounts. To verify whether the account-level variation matters, we first fitted a null model with no predictors and a random intercept. The intraclass correlation coefficient (ICC) was .3, indicating that using a mixed-effects model rather than an OLS regression model is necessary (Cohen, 1988).
11.  $Exp(0.289) - 1 = 0.335$ .
12.  $Exp(-0.922) - 1 = -0.602$ .

## References

- Anderson AA, Brossard D, Scheufele DA, et al. (2014) The “nasty effect”: online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication* 19(3): 373–387.
- Balliet D, Wu J and De Dreu CK (2014) Ingroup favoritism in cooperation: a meta-analysis. *Psychological Bulletin* 140(6): 1556–1581.
- Barnidge M, Ardèvol-Abreu A and Gil De Zúñiga H (2016) Content-expressive behavior and ideological extremity: an examination of the roles of emotional intelligence and information network heterogeneity. *New Media & Society* 20(2): 815–834.
- Barasch A and Berger J (2014) Broadcasting and narrowcasting: how audience size affects what people share. *Journal of Marketing Research* 51(3): 286–299.
- Blackwell C, Birnholtz J and Abbott C (2015) Seeing and being seen: co-situation and impression formation using Grindr, a location-aware gay dating app. *New Media & Society* 17(7): 1117–1136.
- Brewer MB (2001) Ingroup identification and intergroup conflict: when does ingroup love become outgroup hate? In: Ashmore RD, Jussim LJ and Wilder D (eds) *Social Identity, Intergroup Conflict, and Conflict Reduction*. New York: Oxford University Press, pp. 17–41.

- Chen GM and Lu S (2017) Online political discourse: exploring differences in effects of civil and uncivil disagreement in news website comments. *Journal of Broadcasting & Electronic Media* 61(1): 108–125.
- Chen A, Chen K, Zhang J, Meng J and Shen C (2023) When national identity meets conspiracies: the contagion of national identity language in public engagement and discourse about COVID-19 conspiracy theories. *Journal of Computer-Mediated Communication* 28(1):1–12.
- Choi S, Yang JS and Chen W (2018) Longitudinal change of an online political discussion forum: antecedents of discussion network size and evolution. *Journal of Computer-Mediated Communication* 23(5): 260–277.
- Coe K, Kenski K and Rains SA (2014) Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication* 64(4): 658–679.
- Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. London: Routledge.
- Crawford K and Gillespie T (2016) What is a flag for? Social media reporting tools and the vocabulary of complaint. *New Media & Society* 18(3): 410–428.
- Esau K, Friess D and Eilders C (2017) Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet* 9(3): 321–342.
- Evans SK, Pearce KE, Vitak J, et al. (2017) Explicating affordances: a conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication* 22(1): 35–52.
- Eveland WP and Hively MH (2009) Political discussion frequency, network size, and “heterogeneity” of discussion as predictors of political knowledge and participation. *Journal of Communication* 59(2): 205–224.
- Fang K (2022) What is Zimeiti? The commercial logic of content provision on China’s social media platforms. *Chinese Journal of Communication* 15(1): 75–94.
- Festinger L, Pepitone A and Newcomb T (1952) Some consequences of de-individuation in a group. *The Journal of Abnormal and Social Psychology* 47(Suppl. 2): 382–389.
- Fox J and McEwan B (2017) Distinguishing technologies for social interaction: the perceived social affordances of communication channels scale. *Communication Monographs* 84(3): 298–318.
- Gastil J (2008) *Political Communication and Deliberation*. New York: Sage.
- Gibson JJ (1977) The theory of affordances. In: Shaw R and Bransford J (eds) *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale, NJ: Lawrence Erlbaum, pp. 67–82.
- Gil De Zúñiga H and Valenzuela S (2011) The mediating path to a stronger citizenship: online and offline networks, weak ties, and civic engagement. *Communication Research* 38(3): 397–421.
- Gillespie T (2018) *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.
- Gil-Lopez T, Shen C, Benefield GA, et al. (2018) One size fits all: context collapse, self-presentation strategies and language styles on Facebook. *Journal of Computer-Mediated Communication* 23(3): 127–145.
- Guan T, Liu T and Yang Y (2021) Moving away from partisanship: diversifying the motives behind political polarization. *Political Studies Review* 19(4): 656–667.
- Habermas J (1987) *Theory of Communicative Action, Volume Two: Lifeworld and System: A Critique of Functionalist Reason*. Boston, MA: Beacon Press.
- Halpern D and Gibbs J (2013) Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior* 29(3): 1159–1168.
- Han R (2018a) To comply or to resist? The intermediaries’ dilemma. In: Han R (ed.) *Contesting Cyberspace in China*. New York: Columbia University Press, pp. 55–76.

- Han X (2018b) Searching for an online space for feminism? The Chinese feminist group Gender Watch Women's Voice and its changing approaches to online misogyny. *Feminist Media Studies* 18(4): 734–749.
- Hjorth L and Gu K (2012) The place of emplaced visualities: a case study of smartphone visuality and location-based social media in Shanghai, China. *Continuum* 26(5): 699–713.
- Hoffmann M and Heft A (2020) “Here, there and everywhere”: classifying location information in social media data—possibilities and limitations. *Communication Methods and Measures* 14(3): 184–203.
- Jaidka K, Zhou A and Lelkes Y (2019) Brevity is the soul of Twitter: the constraint affordance and political discussion. *Journal of Communication* 69(4): 345–372.
- Jaidka K, Zhou A, Lelkes Y, et al. (2022) Beyond anonymity: network affordances, under deindividuation, improve social media discussion quality. *Journal of Computer-Mediated Communication* 27: 1–3.
- Jia L and Han X (2020) Tracing Weibo (2009–2019): the commercial dissolution of public communication and changing politics. *Internet Histories* 4(3): 304–332.
- John NA and 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.
- Kim B (2020) Effects of social grooming on incivility in COVID-19. *Cyberpsychology, Behavior, and Social Networking* 23(8): 519–525.
- Kim JW, Guess A, Nyhan B, et al. (2021) The distorting prism of social media: how self-selection and exposure to incivility fuel online comment toxicity. *Journal of Communication* 71(6): 922–946.
- Klofstad CA, Sokhey AE and McClurg SD (2013) Disagreeing about disagreement: how conflict in social networks affects political behavior. *American Journal of Political Science* 57(1): 120–134.
- Knustad M and Johansson C (2021) Anonymity and inhibition in newspaper comments. *Information* 12(3): 106.
- Lane DS (2019) Social media design for youth political expression: testing the roles of identifiability and geo-boundedness. *New Media & Society* 22(8): 1394–1413.
- Lee AHY (2021) How the politicization of everyday activities affects the public sphere: the effects of partisan stereotypes on cross-cutting interactions. *Political Communication* 38(5): 499–518.
- Lee FL, Liang H and Tang GK (2019) Online incivility, cyberbalkanization, and the dynamics of opinion polarization during and after a mass protest event. *International Journal of Communication* 13: 20.
- McKune S (2015) “Foreign hostile forces”: the human rights dimension of China's cyber campaigns. In: Lindsay JR, Cheung TM and Reveron DS (eds) *China and Cybersecurity: Espionage, Strategy, and Politics in the Digital Domain*. New York: Oxford University Press, pp. 260–296.
- Mann L, Newton JW and Innes J (1982) A test between deindividuation and emergent norm theories of crowd aggression. *Journal of Personality and Social Psychology* 42(2): 260–272.
- Marchal N (2022) “Be nice or leave me alone”: an intergroup perspective on affective polarization in online political discussions. *Communication Research* 49(3): 376–398.
- Morgan SL and Winship C (2015) *Counterfactuals and Causal Inference*. Cambridge: Cambridge University Press.
- Moy P and Gastil J (2006) Predicting deliberative conversation: the impact of discussion networks, media use, and political cognitions. *Political Communication* 23(4): 443–460.
- Munger K (2017) Tweetment effects on the Tweeted: experimentally reducing racist harassment. *Political Behavior* 39(3): 629–649.

- Mutz DC (2002) Cross-cutting social networks: testing democratic theory in practice. *American Political Science Review* 96(1): 111–126.
- Mutz DC (2006) *Hearing the Other Side: Deliberative Versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Neubaum G and Weeks B (2023) Computer-mediated political expression: a conceptual framework of technological affordances and individual tradeoffs. *Journal of Information Technology & Politics* 20(1): 19–33.
- Oz M, Zheng P and Chen GM (2018) Twitter versus Facebook: comparing incivility, impoliteness, and deliberative attributes. *New Media & Society* 20(9): 3400–3419.
- Pan J and Xu Y (2018) China's ideological spectrum. *The Journal of Politics* 80(1): 254–273.
- Papacharissi Z (2004) Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society* 6(2): 259–283.
- Peng AY (2021) Amplification of regional discrimination on Chinese news portals: an affective critical discourse analysis. *Convergence* 27(5): 1343–1359.
- Penney JW (2016) Chilling effects: online surveillance and Wikipedia use. *Berkeley Technology Law Journal* 31(1): 117–182.
- Perfumi SC, Bagnoli F, Caudek C, et al. (2019) Deindividuation effects on normative and informational social influence within computer-mediated communication. *Computers in Human Behavior* 92: 230–237.
- Peterson E and Muñoz M (2022) “Stick to sports”: evidence from sports media on the origins and consequences of newly politicized attitudes. *Political Communication* 39(4): 454–474.
- Postmes T, Spears R, Sakhel K, et al. (2001) Social influence in computer-mediated communication: the effects of anonymity on group behavior. *Personality and Social Psychology Bulletin* 27(10): 1243–1254.
- Qian Y, Jin Y and Xie Y (2021) Belief in science and attitudes toward COVID-19: a demographic standardization approach to China–US comparison, 2020. *China CDC Weekly* 3(30): 645–649.
- Rains SA, Kenski K, Coe K, et al. (2017) Incivility and political identity on the Internet: intergroup factors as predictors of incivility in discussions of news online. *Journal of Computer-Mediated Communication* 22(4): 163–178.
- Ramsay CR, Matowe L, Grilli R, et al. (2003) Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. *International Journal of Technology Assessment in Health Care* 19(4): 613–623.
- Reicher SD, Spears R and Postmes T (1995) A social identity model of deindividuation phenomena. *European Review of Social Psychology* 6(1): 161–198.
- Rowe I (2015) Civility 2.0: a comparative analysis of incivility in online political discussion. *Information, Communication & Society* 18(2): 121–138.
- Santana AD (2014) Virtuous or vitriolic: the effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice* 8(1): 18–33.
- Schwartz R and Halegoua GR (2015) The spatial self: location-based identity performance on social media. *New Media & Society* 17(10): 1643–1660.
- Scott CR and Rains SA (2020) (Dis) connections in anonymous communication theory: exploring conceptualizations of anonymity in communication research. *Annals of the International Communication Association* 44(4): 385–400.
- Settle JE (2018) *Frenemies: How Social Media Polarizes America*. Cambridge: Cambridge University Press.
- Shen C, Chen A, Luo C, et al. (2020) Using reports of symptoms and diagnoses on social media to predict COVID-19 case counts in mainland china: observational infoveillance study. *Journal of Medical Internet Research* 22(5): e19421.

- SinaTech (2022) The adjusted operating profit margin of Weibo Q2 with a revenue of 3 billion yuan was 32%, a quarter-on-quarter increase. Available at: <https://finance.sina.com.cn/tech/internet/2022-09-01/doc-imizmscv8702747.shtml?cref=cj> (accessed 18 January 2023).
- Sproull L and Kiesler S (1986) Reducing social context cues: electronic mail in organizational communication. *Management Science* 32(11): 1492–1512.
- Su CC, Chan M and Paik S (2022) Telegram and the anti-ELAB movement in Hong Kong: reshaping networked social movements through symbolic participation and spontaneous interaction. *Chinese Journal of Communication* 15(3): 431–448.
- Suler J (2004) The online disinhibition effect. *CyberPsychology & Behavior* 7(3): 321–326.
- Tajfel HE (1978) *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. Cambridge, MA: Academic Press.
- Vraga EK, Thorson K, Kligler-Vilenchik N, et al. (2015) How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior* 45: 281–289.
- Wang J and Sundar SS (2022) Liking versus commenting on online news: effects of expression affordances on political attitudes. *Journal of Computer-Mediated Communication* 27: 1–13.
- Weibo (2022) IP location function update announcement. Available at: [https://weibo.com/1934183965/LqvYeCdBu?ssl\\_rnd=1667059481.8345&type=comment](https://weibo.com/1934183965/LqvYeCdBu?ssl_rnd=1667059481.8345&type=comment) (accessed 18 January 2022).
- Yang G (2009) *The Power of the Internet in China: Citizen Activism Online*. New York: Columbia University Press.
- Yang G (2018) Demobilizing the emotions of online activism in China: a civilizing process. *International Journal of Communication* 12: 21.
- Yang J, Barnidge M and Rojas H (2017) The politics of “unfriending”: user filtration in response to political disagreement on social media. *Computers in Human Behavior* 70: 22–29.
- Yang T and Fang KC (2023) How dark corners collude: a study on an online Chinese alt-right community. *Information, Communication & Society* 26: 441–458.
- Zhang C (2020) Right-wing populism with Chinese characteristics? Identity, otherness and global imaginaries in debating world politics online. *European Journal of International Relations* 26(1): 88–115.
- Zhou A (2021) Causal effects of affordance change on communication behavior: empirical evidence from organizational and leadership social media use. *Telematics and Informatics* 59: 101549.

## Author biographies

Yufan Guo is a master’s student at the School of Journalism and Communication, Tsinghua University. His research interest includes political communication, communication technology, and social change.

Yuhan Li is a doctoral student at the Department of Communication and Media, University of Michigan—Ann Arbor. She is interested in social media, political communication, and computational social science.

Tian Yang is an assistant professor at the School of Journalism and Communication, The Chinese University of Hong Kong. His research interest is at the intersection of digital media, political communication, and computational social science.