When Disagreement Becomes Uncivil on Social Media: The Role of Passive Receiving and Active Expression of Incivility in Influencing Political Polarization

Hsuan-Ting Chen[1] iD **, Yunya Song[2]** iD **, and Jing Guo[3]** iD

Abstract

Exposure to cross-cutting viewpoints may not always play the deliberative role it is supposed to. This study uses both panel survey and social media data to examine how disagreement can trigger incivility, including exposure to and expression of incivility, and further elicit emotions and influence polarization. Results from the two-wave panel survey indicate that cross-cutting exposure has a polarizing effect first through promoting exposure to uncivil messages and expression of uncivil opinions, then through negative emotions. Notably, cross-cutting exposure can indirectly reduce polarization by first encouraging expression of uncivil opinions and then eliciting positive emotions, highlighting the importance of active expression. Analysis of data from the Hong Kong-based discussion forum HKDisc demonstrates that cross-cutting exposure is positively related to exposure to uncivil messages, and exposure to and expression of incivility predict polarization regardless of whether positive or negative emotions are detected in the uncivil content. This study provides empirical evidence of the effects of cross-cutting exposure and incivility on polarization at the individual and collective levels.

[1]The Chinese University of Hong Kong, New Territories, Hong Kong
[2]Hong Kong Baptist University, Kowloon, Hong Kong
[3]Hong Kong Shue Yan University, Hong Kong Island, Hong Kong

Corresponding Author:
Yunya Song, School of Communication, Hong Kong Baptist University, 5 Hereford Road, Kowloon Tong, Kowloon, Hong Kong.
Email: yunyasong@hkbu.edu.hk

Ideally, exposure to cross-cutting viewpoints, a core element of deliberation, should promote diverse opinions and foster respect for different perspectives (Gutmann & Thompson, 1998). However, studies have yielded mixed results on its depolarizing effects, particularly on social media platforms (e.g., Beam et al., 2018; Taber & Lodge, 2006). The rapid growth of social media penetration has led to a global ecosystem of social interaction through daily digital networks with news feeds, comment chains, status updates, and likes. Although it empowers users with various digital affordances, this interaction fosters the extensive spread of uncivil online behaviors due to features like anonymity, missing facial cues in virtual spaces, and the lack of in-person consequences (Papacharissi, 2002). In particular, when social media facilitate exposure to diverse perspectives (Bakshy et al., 2015; Barnidge, 2020), users often witness or become involved in political discussions on contested issues in which commenters flame each other with nonsensical arguments, rude comments, and stereotypes (Hopp & Vargo, 2017; Sun et al., 2021; Theocharis et al., 2016). This study examines the effect of cross-cutting exposure on polarization, as disagreement is likely to trigger incivility on social media. We suggest that the presence of uncivil discourse, particularly in online spaces when encountering opposing viewpoints from outgroups, could serve as one of the mechanisms underlying the inconsistent effects of cross-cutting exposure on the development of deliberative democracy.

Incivility has been shown to have a negative impact on society, including eroding public trust, cultivating toxic culture, exacerbating opinion polarization, and jeopardizing the development of democracy (Anderson et al., 2014). Encountering incivility also wastes people's time and energy in dwelling on others' rudeness and recovering from depression, anger, and fear (Kim & Kim, 2019). However, some scholars suggest the positive potential of online incivility as it can represent resistance against social injustice and stereotypes and mobilize political action (Rains et al., 2017). In addition, deliberative comments may diverge from politeness norms (G. M. Chen, 2017). Such expressive behaviors, even though uncivil, can be deliberative and can induce enthusiasm about political affairs (Kosmidis & Theocharis, 2020).

In this study, we distinguish incivility as *exposure* to uncivil messages and *expressing* uncivil opinions. First, it is possible that people may encounter uncivil messages or comments, and they may as well express their opinions in an uncivil manner after exposure to disagreement. Second, the difference between passive receiving and active expression of incivility has not yet been paid enough attention, but they can induce different emotions (i.e., positive and negative) and affect polarization in different ways (G. M. Chen, 2017). To understand the whole process from cross-cutting exposure to (de)polarization, we propose a serial mediation that cross-cutting exposure can prompt incivility (i.e., exposure to uncivil messages and expressing uncivil opinions), which will trigger emotional arousal (i.e., positive and negative) and further (de)polarize attitude.

This study focuses on Hong Kongers' reactions to the recent wave of emigration from Hong Kong since 2019. This is an emerging and controversial issue in Hong Kong with two opposite stances (support vs. against emigration; see Appendix A for detailed issue context).[1] We adopt two methodological approaches to examine the role of incivility in the effect of cross-cutting exposure on (de)polarization on social media. We first analyze a two-wave panel survey to understand the proposed serial mediation model at the individual level (i.e., the (de)polarization process of cross-cutting exposure through incivility and emotions). Second, to have a sense of what is happening on social media platforms in Hong Kong, we examine how cross-cutting exposure, online incivility, emotions, and polarization at a collective level relate to each other on the Hong Kong Discussion Forum (HKDisc), the most popular discussion forum in Hong Kong.[2] These two methodological approaches provide a more holistic view of the proposed relationships and add to the growing literature on incivility, which has so far mostly relied on experiments by exploring the underlying mechanism and examining the polarization phenomenon on social media (e.g., G. M. Chen & Lu, 2017; Liang & Zhang, 2021).

The (De)Polarizing Effects of Cross-Cutting Exposure

Cross-cutting exposure refers to exposure to political perspectives that conflict with one's existing beliefs (Mutz, 2002). Deliberative theorists suggest that cross-cutting exposure drives individuals to learn from opposing views and increases their political tolerance (Mutz, 2002). Cross-cutting exposure should also stimulate news elaboration and deliberative discussion (Delli Carpini et al., 2004). After talking with others from various backgrounds, users could become less extreme and rigid in their views as a result of managing a diverse social network. Thus, cross-cutting exposure is expected to neutralize one's political stance and attenuate political polarization (Wojcieszak & Price, 2009). Beam et al. (2018) found that social media news use enables people to consume both pro- and counter-attitudinal information, and individuals depolarize over time after consuming counter-attitudinal information. Similarly, Wojcieszak and Garrett (2018) observed that exposure to a larger number of counter-attitudinal news articles was negatively related to all indicators of affective polarization.

However, cross-cutting exposure may not always play a positive role. Some scholars have found that rather than reducing extreme attitudes, it can actually intensify extremism (Bail et al., 2018). When people encounter messages that contradict their views, they tend to counterargue them through motivated reasoning and confirmation bias, which reinforce their preexisting beliefs (Taber & Lodge, 2006; Wojcieszak & Price, 2010). They may also filter out conflicting viewpoints and refuse to process them (H.-T. Chen et al., 2022; Guo & H.-T. Chen, 2022; Lin et al., 2023). Thus, cross-cutting exposure may backfire, contributing to increased political polarization. In this study, we focus on social media platforms because they facilitate exposure to cross-cutting viewpoints (Bakshy et al., 2015) and are forums where political issues are frequently discussed.

To understand polarization, most previous studies have examined two distinct forms: ideological and affective polarization (Kubin & Sikorski, 2021). Ideological polarization refers to the divergence of political opinions, attitudes, and stances of politically rival groups (Dalton, 1987), while affective polarization focuses on how people like in-group allies and dislike out-group opponents (Iyengar et al., 2019). In this study, we focus on ideological polarization given that we can take advantage of the two-wave panel survey and social media data to understand this type of polarization from a more holistic view by examining the data at individual and collective levels. At the individual level, polarization is captured by the number of positions in the two extremes of a continuum (i.e., attitude extremity; Mason, 2015).[3] At the collective level, it is the extent to which opinions are diverse (i.e., dispersion) or move toward separate camps (i.e., bimodality; DiMaggio et al., 1996). Analyzing survey results allows the examination of the divide of issue positions (i.e., against versus supporting emigration), while analyzing social media data illustrates the dispersion of opinions about the issue.

The mixed findings in prior research on the relationship between cross-cutting exposure and polarization could be due to different underlying mechanisms in the relationship. We examine the role of incivility and the emotional arousal it engenders in the relationship because exposure to counter-attitudinal information often triggers heated discussion, including uncivil messages on social media where users can easily read the messages or express their opinions.

Cross-Cutting Exposure and Online Incivility

Online incivility commonly refers to foul language used in online discourse (F. L. F. Lee et al., 2019). Scholars have adopted personal-level and public-level approaches to study incivility (Muddiman, 2017). In the public-level approach, online incivility is viewed as verbalized threats to democracy and personal freedom, such as stereotypes, conspiracy theories, populism, and hate speech (Papacharissi, 2004). In the personal-level approach, online incivility goes against politeness norms. In line with previous online incivility studies from the personal-level approach (e.g., G. M. Chen & Lu, 2017; Gervais, 2014; Y. Song et al., 2022), our conceptualization of incivility is built on the premise of politeness theory (Mutz, 2015) and operationalized as foul language used in online discourse (F. L. F. Lee et al., 2019). Specifically, we take the personal-level approach in conceptualizing online incivility as name-calling, aspersion, vulgarity, swearing, and mockery that make individuals or actions seem radical, immoral, or corrupt during online discussions (Muddiman, 2017). This line of research commonly centers on the predictors and effects of uncivil discourse that offend politeness norms. We argue that when investigating the antecedents and outcomes of incivility, it is important to examine exposure to and expression of incivility separately. Although both concepts pertain to uncivil discourse in general, they are conceptually different as one relates to receiver effects and the other to sender effects (Liang & Ng, 2023; Pingree, 2007).

Previous studies have examined either passive exposure to incivility (e.g., Anderson et al., 2014; Stroud, 2017), or active expression of incivility (e.g., Oz et al., 2018).

For instance, Anderson et al. (2014) found that when exposed to uncivil blog comments, those who do not support nanotechnology are more likely to perceive the technology as risky than those who are exposed to civil comments. People are also more likely to respond negatively when they are exposed to online incivility that is directed at them or their views (Phillips & Smith, 2004). Focusing on expression effects, Liang and Ng (2023) suggest that expressing uncivil disagreement increases anger and perceptions of incivility.

In this study, we examine both concepts simultaneously to understand their differential effects on polarization. Exposure to uncivil messages is a passive receiving of incivility, which refers to the frequency and amount of access to personal attacks, vilification, accusation, name-calling, and disparagement online (Gervais, 2014). By contrast, the expression of incivility is a proactive behavior in which individuals themselves use uncivil words and actions during political discussions. Hmielowski et al. (2014) operationalized online uncivil expression as flaming behavior when faced with a direct challenge to one's beliefs.

Online incivility usually happens in discussions on political events, such as elections, and controversial issues, such as immigration or racism (Hopp & Vargo, 2017). Controversy entails diverse viewpoints and the potential for conflict. Nevertheless, cross-cutting exposure could provide alternative views, helping individuals to understand the opposite dimension of public opinion and leave their echo chambers. Uncivil content will inevitably emerge during controversial discussions with heterogeneous perspectives. It could be profoundly uncomfortable as people of different backgrounds and values participate (Schudson, 1997). Thus, cross-cutting exposure could increase the likelihood of exposure to uncivil messages.

In addition, the potential conflicts that come from cross-cutting exposure could also motivate individuals to express themselves in an uncivil way. When cross-cutting exposure involves heterogeneous others with opposite political views, there is less mutual trust, understanding, and empathy compared with attitude-consistent exposure (H.-T. Chen & Lin, 2021; Chan et al., 2021). Thus, people are more likely to use uncivil manners to counter oppositional groups in conflict resolution. In addition to exposure to uncivil messages, cross-cutting exposure should foster uncivil expression. We propose the first hypothesis:

H1: Cross-cutting exposure is positively related to exposure to uncivil messages (H1a) and expressing uncivil opinions (H1b).

Emotional Responses to Incivility

Emotional responses are elicited by individuals' subjective evaluations and understanding of their surroundings (Omdahl, 2014). They play an important role in conduiting the effect of uncivil disagreement on political outcomes (G. M. Chen, 2017). Most prior studies on incivility have focused on negative emotional responses as the outcome of exposure to incivility because aggressive and hostile language is aversive and likely to induce moral anger and discomfort (e.g., Kim & Kim, 2019). However,

it is equally important to acknowledge the role of positive emotions when studying incivility for two theoretical reasons. First, in addition to passively receiving uncivil messages, people can actively express themselves in an uncivil manner when encountering disagreement. The sender effects could be different from the exposure effects. Second, negative and positive emotions are two separate experiences and not mutually exclusive (G. M. Chen, 2017). For instance, the presence of anger does not suggest the absence of enthusiasm. The two different experiences may co-exist. Thus, it is crucial to explore how exposure to uncivil messages and the expression of uncivil opinions can elicit two types of emotions and further result in distinct political consequences (Marcus et al., 2000).

Negative Emotional Responses to Incivility

After exposure to uncivil messages online, people are likely to evaluate the uncivil content and consider the uncivil action as a violation of social norms. Prior studies have consistently demonstrated that exposure to uncivil messages triggers negative emotions, especially when the messages are disagreeing (e.g., G. M. Chen & Lu, 2017). When people perceive incivility as bad and harmful, negative emotions like anger, disgust, and contempt are likely to be induced (M. Y. Wang & Silva, 2018). Aggressive content is also likely to affect the receiver's inner state and satisfaction with the online discourse, yielding an experience of negative emotions (Gervais, 2014).

Emotional arousal often occurs in the interactive relational context in addition to information exposure. When exposure to political disagreement leads to cognitive dissonance (Festinger, 1957), people may try to relieve their cognitive discomfort by rationalizing what happened, often by blaming out-group attackers and defending in-group legitimacy in ways that may involve uncivil interactions. Foul language, hate speech, or name-calling as a way of expressing oneself in a highly emotionally aroused manner for defensive purposes is likely to trigger negative emotions like anger and resentment during the uncivil interactions. Thus, this study proposes that both exposure to uncivil messages and expressing uncivil opinions in a cross-cutting exposure context are likely to elicit negative emotions.

H2: Exposure to uncivil messages (H2a) and expressing uncivil opinions (H2b) are positively related to negative emotions.

Positive Emotional Responses to Incivility

Positive emotions like hope and enthusiasm have been underexplored compared to negative emotions and have rarely been examined simultaneously with negative emotions in previous studies on incivility (e.g., Kim & Kim, 2019). Some researchers suggest that exposure to incivility may induce positive emotions when the uncivil messages are like-minded because uncivil language makes the content less formal and more entertaining to read (e.g., Berry & Sobieraj, 2014). Like-minded messages can

reaffirm people's political stance and increase group solidarity (Brooks & Geer, 2007). For instance, Kosmidis and Theocharis (2020) found that exposure to uncivil tweets increases enthusiasm if the main argument is in line with the dispositions of the audience. However, the situation is less apparent when the uncivil messages are *unlikeminded*, a setting in our study. Theoretically, exposure to disagreeing uncivil messages should mitigate positive emotions. G. M. Chen (2017) examined whether exposure to uncivil disagreement can lead to a burst in negative emotions and a deflation of positive emotions but found no significant relationship for positive emotions. The relationship between exposure to incivility, especially in a counter-attitudinal context, and positive emotions is not clear and needs to be further examined. Thus, we raise a research question:

RQ1: Does exposure to uncivil messages relate to positive emotions?

When it comes to expressing one's opinion uncivilly, we assume that expressing uncivil opinions in a cross-cutting context may not only elicit negative emotions, as discussed above, but also trigger positive emotions based on two theoretical reasons. First, expressing uncivil opinions is a way to release strong aggressive feelings, which could reduce dissatisfaction toward either political issues or the uncivil interaction. In addition, expressive behaviors can enhance political efficacy and knowledge because expressing opinions and having political conversations with others help people be confident in their competency to talk about and understand political issues (Chan et al., 2017).

Second, cross-cutting exposure may trigger a defensive mechanism that makes people express themselves in uncivil ways against the out-group (G. M. Chen, 2017). Indeed, the defensive mechanism would trigger negative emotions, but uncivil expression that protects the in-group and attacks the out-group could also provoke positive emotions, such as enthusiasm. Expressing uncivil opinions to correct or counterargue oppositional viewpoints can enhance group solidarity and the feeling of validation in one's political stance. Expressing one's opinion uncivilly, therefore, could prompt positive emotions. Following this reasoning, we propose the following hypothesis:

H3: Expressing uncivil opinions is positively related to positive emotions.

Emotional Responses and Polarization

Emotions affect how people process information and make decisions as they provide cues and guide people on how to feel about and react to the information they encounter (Marcus et al., 2000). Thus, emotions have been found to play an important role in influencing political perceptions, behaviors, and attitudes (H. H. Lee & Kwak, 2014). More importantly, positive and negative emotions are likely to have different effects on political outcomes as they have differential effects on reward-seeking or danger-averting behaviors (Marcus et al., 2000).

Negative emotions, such as anger and disgust, have been found to mobilize political participation (Valentino et al., 2011); however, scholars have expressed concerns about whether this type of participation is healthy for the development of democracy as it is prompted by citizens' withdrawal from participating in the deliberative process. For instance, anxiety promotes attention and actions to prevent threats (e.g., unwanted policies) from occurring, while anger leads people to protect their beliefs and confront adversaries (Marcus et al., 2000). This defensive mechanism, triggered by threats or attacks imposed by the oppositional out-group, is likely to make people biasedly evaluate the in-group and out-group and become more entrenched in their existing opinions, exacerbating political polarization (Kim & Kim, 2019).

On the other hand, positive emotions, such as hope and enthusiasm, are driving forces for effective learning processes (Pekrun, 1992). Passionate and competent feelings toward political information should lead people to pay more attention to the issue and encourage more careful deliberative elaboration instead of driving attention away or discouraging reasoning on the issue. Thus, positive emotions should produce a "broad, flexible cognitive organization and ability to integrate diverse materials" (Isen, 1990, p. 89), and widen the array of thoughts that come to mind (Fredrickson & Branigan, 2001), which should make political attitudes less extreme. In an experiment, McLaughlin et al. (2020) examined participants' enthusiasm for an in-group or an out-group candidate and its relationship with polarization. Their participants became less polarized only when they were enthusiastic toward the unlike-minded source (i.e., the out-group candidate). Accordingly, in our study, which is a cross-cutting exposure context, positive emotional response to incivility should diminish polarization. We propose the following hypotheses:

H4a: Negative emotions are positively related to ideological polarization.
H4b: Positive emotions are negatively related to ideological polarization.

Serial Mediating Relationships

All the direct relationships discussed above will be examined using social media data and two-wave panel data. In addition, we take advantage of the two-wave panel design to understand the whole process. We propose a serial mediation model (Figure 1) that builds on the direct relationships and bridges the pathways of the indirect effects of cross-cutting exposure to polarization through incivility (i.e., exposure to and expression of incivility) and emotions (i.e., positive and negative emotions). A set of hypotheses and a research question (due to the uncertain relationship between uncivil exposure and positive emotions as raised in RQ1) are presented about the indirect effects in the model:

H5a: The indirect effect of exposure to cross-cutting viewpoints on polarization is serially mediated first through exposure to uncivil messages, then through negative emotions.
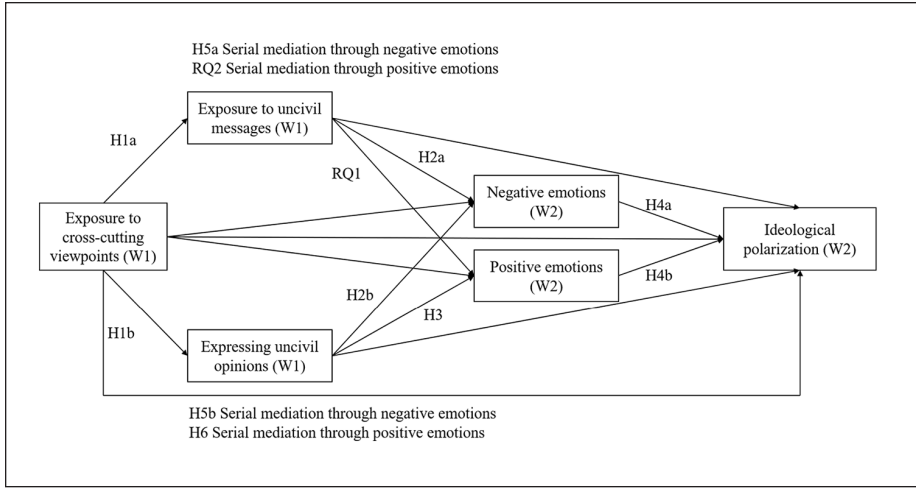
Figure 1. The proposed serial mediation model: The indirect effect of exposure to cross-cutting viewpoints on political polarization first through exposure to uncivil messages or expressing uncivil opinions, and then through negative or positive emotions.
Note. Control variables and Wave 1 measures of negative emotions, positive emotions, and ideological polarization are controlled but not presented in the figure to enhance readability.

H5b: The indirect effect of exposure to cross-cutting viewpoints on polarization is serially mediated first through expressing uncivil opinions, then through negative emotions.

RQ2: Is there an indirect effect of exposure to cross-cutting viewpoints on depolarization that is serially mediated through exposure to uncivil messages and positive emotions?

H6: The indirect effect of exposure to cross-cutting viewpoints on depolarization is serially mediated first through expressing uncivil opinions, then through positive emotions.

Study 1: Two-Wave Panel Survey

Method

Data. Data were drawn from a two-wave online panel study conducted in Hong Kong by Dynata, an online survey panel company. The first wave of data collection (W1) was conducted from March 17 to 29, 2021, with a total of 1787 participants. The second wave (W2) took place from May 21 to June 21, 2021, with 1238 of the original respondents completing the survey questionnaire, yielding a good retention rate of 69.28%. Participants were Hong Kong citizens aged above 18 years. To proportionally represent the Hong Kong population, Dynata employed stratified quota sampling with the age, gender, income, and education quotas specified so that the sample would

match the distribution of these demographic variables as reported by the Hong Kong Census (See Appendix B).

Measures

Exposure to Cross-Cutting Viewpoints. Adopted from previous studies (Barnidge, 2017), the measure of exposure to cross-cutting viewpoints asked respondents how often in the past month they had encountered political information on social media related to the Hong Kong emigration issue that (a) disagreed with your political views, (b) was critical of viewpoints you support, and (c) was supportive of viewpoints you oppose (1 = *never* to 5 = *always*). The 3 items were averaged to form an index of exposure to cross-cutting viewpoints (W1: α = .88, mean [*M*] = 2.50, standard deviation [*SD*] = 1.29)

Exposure to Uncivil Messages. Exposure to uncivil messages was measured by asking respondents how often in the past month (1 = *never* to 5 = *always*) they had been exposed to the following types of messages when encountering cross-cutting political information on social media related to the Hong Kong emigration issue: (a) aspersion, (b) negative character exaggerations through spin, (c) lying accusation, (d) vulgarity, and (e) pejorative speech. The 5 items were averaged to form an index of exposure to uncivil messages (W1: α = .96, *M* = 2.15, *SD* = 1.35; Gervais, 2014).

Expressing Uncivil Opinions. Expressing uncivil opinions was measured by asking respondents how often in the past month (1 = *never* to 5 = *always*) they had expressed their opinions with the following types of messages when encountering cross-cutting political information on social media related to the Hong Kong emigration issue: including (a) aspersion, (b) negative character exaggerations through spin, (c) lying accusation, (d) vulgarity, and (e) pejorative speech. The 5 items were averaged to form an index of expressing uncivil opinions (W1: α = .96, *M* = 1.46, *SD* = 0.81).

Positive and Negative Emotions. Respondents were asked how often in the past month (1 = *never* to 5 = *always*) they had felt the following emotions when encountering cross-cutting uncivil messages related to the Hong Kong emigration issue: (a) anger, (b) anxiety, (c) fear, (d) disgust, (e) hope, and (f) enthusiasm. The first 4 items were grouped as negative emotions (W1: α = .90, *M* = 2.21, *SD* = 1.10; W2: α = .91; *M* = 2.06, *SD* = 1.06) and the last 2 items were grouped as positive emotions (W1: Spearman-Brown Coefficient = .94, *M* = 1.75, *SD* = 0.98; W2: Spearman-Brown Coefficient = .94, *M* = 1.70, *SD* = 0.97; Dillard & Shen, 2007). A factor analysis showed that the items constructed two clean factors (eigenvalues = 2.91 and 2.22 respectively).

Polarization. Issue-specific ideological polarization was measured by three paired statements related to the Hong Kong emigration issue. Respondents were asked to indicate which statement matched their opinion better on a 5-point scale of 1 = *against emigration position* (e.g., "Emigrating to other countries does not help the current situation in Hong Kong. We should stay here to fight for the future of this city"),

3 = *neutral*, and 5 = *support emigration position* (e.g., "The current situation in Hong Kong will never change. If we cannot have hope, it is understandable to seek hope in other places"; W1: α = .81, *M* = 3.08, *SD* = 1.06; W2: α = .81, *M* = 3.08, *SD* = 1.02. See Appendix C for all the statements). The scale was folded such that higher values indicated greater polarization (range 1 to 3; W1: *M* = 2.01, *SD* = 0.61; W2: *M* = 1.95, *SD* = 0.60).

Control Variables. We controlled demographic variables, including age (*M* = 38.16, *SD* = 11.42), gender (male = 47.2%), education (*Mdn* = college, *SD* = 1.36), and monthly income (*Mdn* = HK$40,000 to under $49,999, *SD* = 2.27). We also controlled news media use (*M* = 5.70, *SD* = 1.78), political interest (W1: *M* = 3.16, *SD* = 1.01), and political efficacy (W1: α = .75, *M* = 2.64, *SD* = 0.85).

Statistical Analysis. Taking advantage of the two-wave panel survey design, we conducted structural equation modeling using Mplus with a panel-lagged and autoregressive analytic approach to examine the serial mediation model. This approach assesses the impact of each independent variable on the change of each outcome variable. This approach regressed each Wave 2 variable on its corresponding Wave 1 variables to isolate its effect over time (Gil de Zúñiga et al., 2014). Given that receiving and expressing incivility are both immediate responses that are possible after cross-cutting exposure, cross-cutting exposure, exposure to uncivil messages, and expressing uncivil opinions are variables from Wave 1. Emotional arousal and different types of polarization are variables from Wave 2 to capture the effect of cross-cutting exposure and incivility on social media. Before testing the model, we created a residualized covariance matrix by regressing all the main variables in the proposed model on the control variables and the auto-regressive terms (i.e., emotions and polarization in W1; see Appendix D for the correlations between the variables).

Study 1 Results

The validity of the proposed model depicted in Figure 1 was tested. Because the proposed model is a saturated model, the model fit is not provided first. Then, we released the insignificant paths to free degrees of freedom for model fit calculation. As shown in Figure 2, the relationships between cross-cutting exposure and exposure to uncivil messages (H1a: *B* = .387, *p* < .001) and between cross-cutting exposure and expressing uncivil opinions (H1b: *B* = .149, *p* < .001) are significant, supporting H1. Exposure to uncivil messages (H2a: *B* = .132, *p* < .001) and expressing uncivil opinions (H2b: *B* = .095, *p* < .01) are significantly related to negative emotions, supporting H2. Expressing uncivil opinions is significantly correlated to positive emotions (H3: *B* = .208, *p* < .001), but exposure to uncivil messages is not (RQ1: *B* = .056). Negative emotion is positively related to polarization (H4a: *B* = .077, *p* < .05), while positive emotion is negatively related to polarization (H4b: *B* = -.123, *p* < .001). H4 is supported.

After examining the saturated model, we released the insignificant paths. The results show a very good fit for the model based on the criteria recommended by Hu and
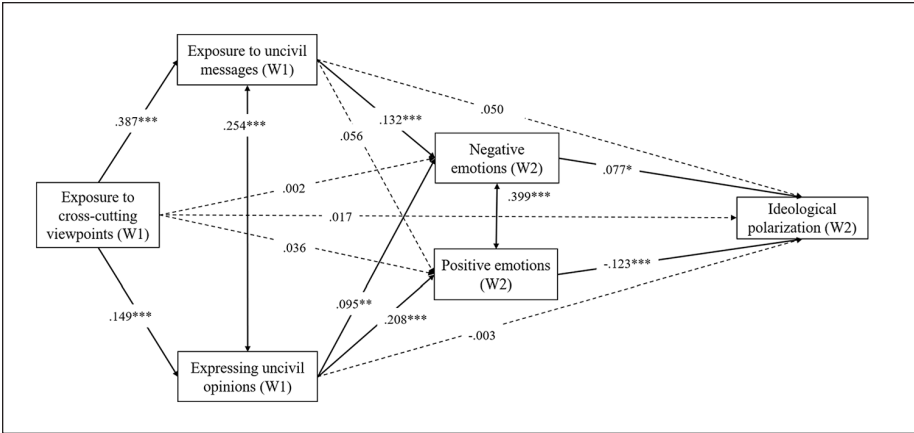
Figure 2. The saturated serial mediation model.
Note. Sample size = 1,234. Path entries are standardized coefficients. Control variables and Wave 1 measures of negative emotions, positive emotions, and ideological polarization are controlled but not presented in the figure to enhance readability. The variables included in this analysis accounted for 15.0% of the variance in exposure to uncivil messages, 2.2% in expressing uncivil messages, 3.4% in negative emotions, 5.8% in positive emotions, and 1.6% in ideological polarization.
*p < .05; **p < .01; ***p < .001.

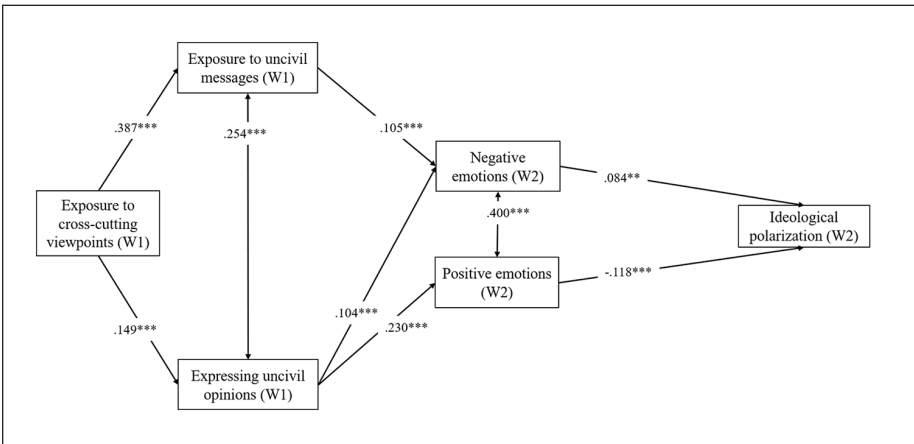**Figure 3.** The revised serial mediation model.
Note. Sample size = 1,234. Goodness of fit: χ^2 = 11.60, *df* = 6, *p* = .07; CFI = .991, TLI = .978, RMSEA = .028, and SRMR = .021. Path entries are standardized coefficients. Control variables and Wave 1 measures of negative emotions, positive emotions, ideological polarization and affective polarization are controlled but not presented in the figure to enhance readability. The variables included in this analysis accounted for 15.0% of the variance in exposure to uncivil messages, 2.2% in expressing uncivil messages, 2.8% in negative emotions, 5.3% in positive emotions, and 1.3% in ideological polarization.
**p < .01; ***p < .001.

Bentler (1999). The chi-square test of model fit is 11.60 with $p = .07$ and $df = 6$, CFI = 99.1, TLI = .978, RMSEA = .028, and SRMR = .021. As shown in Figure 3, the significant paths stay similar. The analysis also identified three mediating relationships between cross-cutting exposure and polarization: (1) cross-cutting exposure can lead to polarization first through prompting exposure to uncivil messages, then through negative emotions (H5a: $\beta = .003$, $p < .05$); (2) cross-cutting exposure can lead to polarization first through expressing uncivil opinions, then through negative emotions (H5b: $\beta = .001$, $p < .05$); and (3) cross-cutting exposure can indirectly diminish polarization first through actively expressing uncivil opinions, then through arousing positive emotions (H6: $\beta = -.004$, $p < .01$).[4]

Study 1 Discussion

The results show that cross-cutting exposure can lead to higher levels of uncivil exposure and expression. However, uncivil exposure, as low-effort passive acceptance, only generates negative emotions that in turn enhance polarization, while uncivil expression not only elicits negative emotions but may also arouse positive emotions that can mitigate political polarization.

There are some implications of these findings. The first is a warning that the hostile atmosphere online could be a key element in exacerbating social fracture and polarization. Cross-cutting exposure triggers incivility in the form of either exposure or expression, which can indirectly lead to polarization because it can elicit negative emotions that are non-deliberative in nature.

Second, the findings highlight the differential effect between exposure to and expression of incivility and the unique role of uncivil expression and positive emotions in mediating the depolarizing effect of cross-cutting exposure. Expressive behavior is considered a high-effort affordance on social media (J. Wang & Sundar, 2022). Expression against disagreeing others could help relieve stress, ease contradiction, and settle conflicts. Thus, through counterarguing or challenging others, even if in an uncivil manner, people may reaffirm their positions and be more involved in the issues.

Although uncivil expression could set off negative emotions such as anger and hostility, it could also trigger positive emotions such as hope and enthusiasm. This finding highlights the potentially positive role of incivility in deliberation as it may facilitate deeper issue learning, mobilize deliberative thinking, and attenuate polarization.

Study 2: Social Media and Computational Content Analytics

Method

Data Collection. Using a list of keywords including "移民" (emigration), "移居" (migration), "移英" (emigration to UK), "移加" (emigration to Canada), "移澳" (emigration

to Australia), "移美" (emigration to US), "BNO" (British National Overseas), "逃出香港/離開香港/逃離香港/離港" (escaping from Hong Kong/departing from Hong Kong/fleeing Hong Kong/leaving Hong Kong), and "北上" (moving north), we collected posts and the corresponding comments related to the emigration issue from HKDisc. At the post level, posting time, user screen name, user ID, and post content were collected. At the comment level, posting time, user screen name, user ID, and comment content were collected. A total of 4,883 posts and 135,962 corresponding comments published between March 2019 and April 2022 were retrieved. To capture the sharing behavior of forum users discussing emigration, the dataset was filtered to retain only those which contain an embedded link (URL) to web content. These links are the digital media items shared by the forum users, including forum posts, news articles, videos, and other content.

Network Analysis. First, we constructed a bipartite network that comprises two node classes (i.e., users and URLs) by creating an edge whenever a forum user shares one URL. To measure the degree of users' preference for sharing the same URL, we used Newman's collaboration model for weight calculation (Newman et al., 2001):

$$w_{u,v} = \sum_k \frac{\delta_u^k - \delta_v^k}{d_k - 1}$$

where u and v represent user nodes and k represents a URL node. The value d_k is the degree of k. δ_u^k is 1 if u has a connection to k or 0 otherwise in the bipartite network.

Second, we used Newman et al.'s (2001) weighted projection method to transform the bipartite network to a unipartite network of only user nodes where edges between nodes represent the connection of users sharing the same URL. The projection enables us to focus the analysis on the relationships between forum users and their collective patterns of sharing behavior. The linkages reveal that users share the same URL and may thus share similar views. To avoid network fragmentation issues, we retained the giant component of the unipartite network for analysis and excluded other components. This giant component consists of around a third of all user nodes (427 out of 1,310 user nodes) and roughly 99% of all edges (4,975 out of 5,020 edges) in the unipartite network of user nodes.

Third, we used the Louvain method of community detection to identify clusters of densely interconnected nodes in the network (Aynaud et al., 2013), which, in this context, represent groups of users who share similar URLs.

Content Analysis. We used structural topic modeling (STM; Roberts et al., 2014), an unsupervised machine learning approach based on statistical algorithms, to identify the most prevalent topics among users in the co-sharing network. Extending the tradition of probabilistic topic models like Latent Dirichlet Allocation (Blei et al., 2003), STM assumes that each document consists of a mixture of topics, and each topic is represented by a list of "top words" with high semantic coherence scores. We aggregated each user's historical posts and comments as a corpus for model construction.

The Python natural language processing package spaCy was utilized for Cantonese word segmentation. Stop words, punctuation, URLs, pure digit numbers, and special symbols were removed from the texts. Words with extremely high or low frequencies in the corpus were also excluded. After comparing models with numbers of topics ranging from 10 to 50, we finally set 20 as the number of topics. Each topic was manually labeled based on the 30 most-probable words. After removing one meaningless topic, 19 topics were obtained. Appendix F details the topic modeling results.

Building on the existing lists of uncivil Cantonese words from previous studies (F. L. F. Lee et al., 2019; Y. Song et al., 2022) and online web sources (e.g., Encyclopedia of Virtual Communities in Hong Kong), we constructed a dictionary of uncivil expressions, defined as language that is crude, vulgar, offensive, or otherwise hurtful in nature. Both uncivil cyber language and daily expressions were considered. Two coders went through the compiled list and only the items agreed upon by both were retained, resulting in a list of 807 uncivil expressions.

Measures

Exposure to Cross-Cutting Viewpoints. For each user's ego network in the co-sharing, we calculated the proportion of out-group edge weights (W. Chen et al., 2021).

Exposure to Uncivil Messages. Likewise, we used the incivility dictionary to calculate exposure to explicit swearing. (1) If a post's text score for explicit swearing was not zero, we considered it an uncivil post message. All users who commented on this post are considered to have been exposed to uncivil messages once. (2) In a post, a commenter is also considered to have been exposed to uncivil messages if there is at least one comment with a non-zero swearing score that was published before the user's comment. We defined exposure to uncivil messages by calculating the number of times each user was exposed to uncivil messages divided by the sum of the user's historical count of both posts and comments.

Expressing Uncivil Opinions. We used the incivility dictionary built in the previous step to calculate scores for the use of explicit swearing. For each user, we calculated the mean score for the use of explicit swearing based on all their historical posts and comments.

Positive and Negative Emotions. We used TextMind, a Chinese linguistic psychological text analysis system, to calculate the scores for positive and negative emotions in the expression.[5] Similar to Linguistic Inquiry and Word Count (LIWC), TextMind assigns preferences and degrees to different psychological categories in the text (Zhao et al., 2016). We would like to highlight that the emotion from the social media data is about the features of the expression (i.e., sentiments), which is conceptually and operationally different from what we measured in the Study 1 survey about emotional arousal.

Polarization. We utilized the results from community detection and topic modeling to calculate the polarization score. The community detection analysis described above
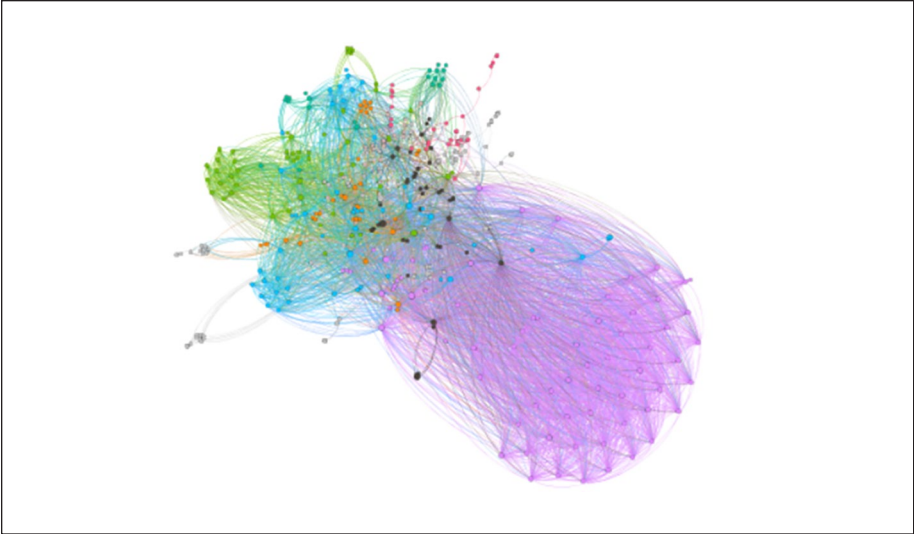
Figure 4. Users' co-sharing network.
Note. Node size controlled by node degree. Layout of this network is ForceAltas2. Only the top 8 largest clusters are colored: purple (17.10% of the total nodes), light green (12.88%), blue (11.94%), black (10.77%), orange (7.96%), pink (6.79%), green (4.92%), and beige (4.68%). Visualization was plotted by Gaphi 0.9.2.

enables us to distinguish the in-group and out-group nodes and edges. In-group edges connect users only within the same community, while out-group edges connect users beyond the same community (De Francisci Morales et al., 2021). We calculated the cosine similarity between each user's topic proportion distribution and the average topic proportion distribution of other in-group members to capture the distance of opinion between users. We also calculated the cosine similarity between the user's and the out-group member's topic proportion distribution. The polarization ratio was defined as in-group similarity divided by the sum of in- and out-group similarity. The larger the ratio, the greater the in-group topic homogeneity, and thus the stronger the echo-chamber effect of the community.

Study 2 Results

Before testing the hypotheses and research question, we first examined the polarization phenomenon on HKDisc. Figure 4 displays a visualization of the co-sharing network. The community detection analysis identified 16 communities. The modularity of this network is .58 (between .3 and .7), indicating that the community assignment is significantly better than a random assignment (Newman & Girvan, 2004). We observed a polarized network structure in which the largest community (in purple) dominated on one side of the network, whereas the other communities were on the other side.
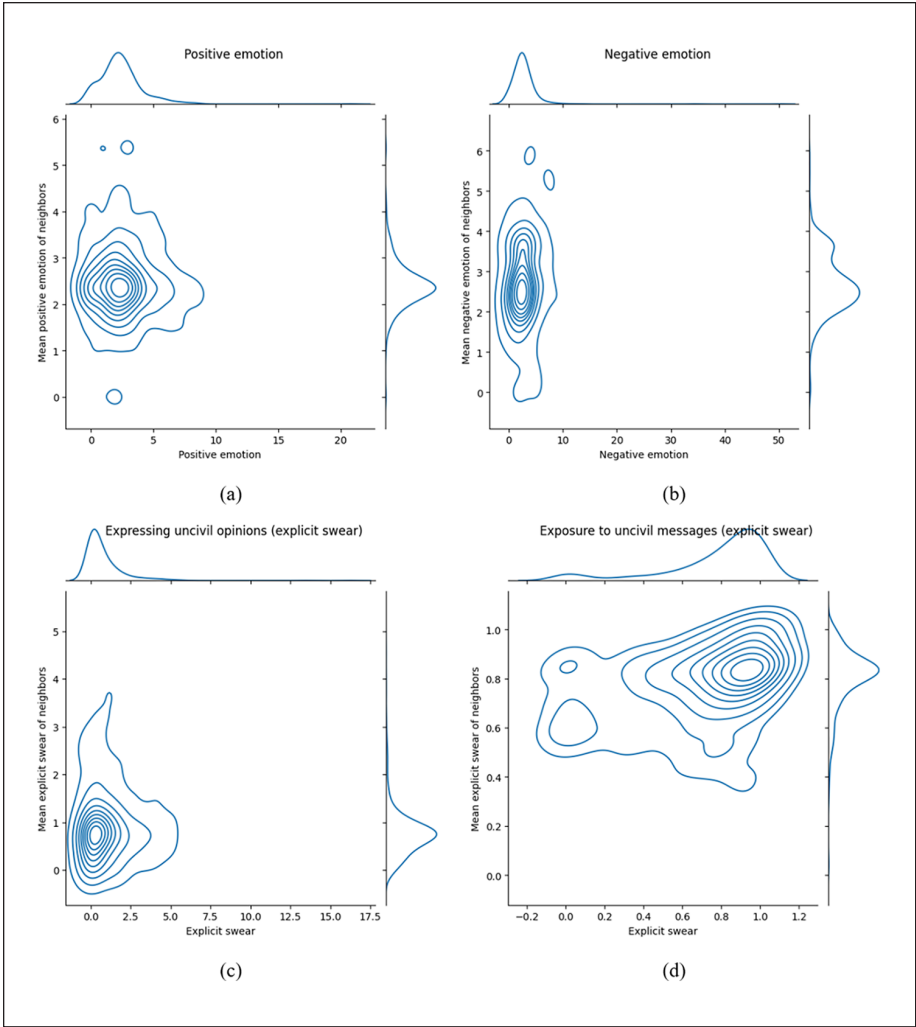
Figure 5. Correlations between users and their neighbors in terms of (a) positive emotion, (b) negative emotion, (c) expressing uncivil opinions, and (d) exposure to uncivil messages. *Note.* For each subplot, the *x*-axis refers to the user and the *y*-axis refers to the user's neighboring node. A denser area in the plots indicates a greater density of users correlated with their neighboring nodes in the co-sharing network.

We then analyzed users' homophily in clusters by calculating the correlation between users and their neighboring nodes in terms of expression of positive emotion, negative emotion, and uncivil opinions and exposure to uncivil messages (Figure 5). The results show that highly-connected users tend to be surrounded by neighbors with

Table 1. Regression Models for Predicting Ideological Polarization.

	DV: Polarization		
IVs	Coef	Std	p-Value
Constant	1.2900	0.034	.000
Cross-cutting exposure	0.0127	0.024	.592
Expressing uncivil opinions	0.0210**	0.007	.002
Exposure to uncivil messages	0.1521***	0.035	.000
Positive emotions	0.0096	0.005	.060
Negative emotions	−0.0020	0.003	.484

Note. N = 426. R^2 = .086.
**$p < .01$. ***$p < .001$.

similar scores in all these variables. This finding is in line with users' homophily in echo chambers where "birds of a feather flock together."

The results of Pearson's correlation analysis show that cross-cutting exposure is positively correlated with exposure to uncivil messages ($r = .177, p < .001$), supporting H1a, but not with expressing uncivil opinions ($r = .057$), rejecting H1b. Although not proposed in the hypothesis, there was a significant relationship between exposure to uncivil messages and expressing uncivil opinions ($r = .170, p < .001$).

Exposure to uncivil messages is not significantly related to negative emotions ($r = .023$). H2a is not supported. Uncivil exposure is also not significantly related to positive emotions (RQ1: $r = −.022$). Expressing uncivil opinions is positively related to negative emotion ($r = .182, p < .001$), supporting H2b, but it is not related to positive emotions (H3: $r = −.034$).

We also found that neither negative emotions (H4a: $r = .008$) nor positive emotions (H4b: $r = .078$) are significantly related to polarization. However, both exposure to uncivil messages ($r = .239, p < .001$) and expression of uncivil opinions ($r = .180, p < .001$) are significantly related to polarization (see Appendix G for the correlation table).

We further conducted OLS regression to test the effects of cross-cutting exposure, exposure to uncivil messages, expression of uncivil opinions, and expression of emotions on polarization. Table 1 presents the results. Only exposure to ($B = .152, p < .001$) and expression of uncivil opinions ($B = .021, p < .01$) significantly predicted polarization.

Study 2 Discussion

Results from the social media data provide a more worrisome picture for the development of a healthy democracy as no depolarizing effects were detected. Cross-cutting exposure is not related to expressing uncivil opinions, but it is related to exposure to uncivil messages. However, exposure to and expression of incivility are related to each other, and both predict polarization. This implies that on a collective level, cross-cut-

ting exposure on social media may indirectly polarize public opinions through widespread nasty talk online.

Cross-cutting exposure can expose individuals to unfavorable opinions and interrupt their experience in an echo chamber; however, it does not directly cause them to express uncivil opinions unless they encounter uncivil messages through this exposure. This finding is different from the results of the two-wave panel survey. Individuals might believe they would respond to or have responded to counter-attitudinal information in an uncivil manner, but in reality, various macro-level factors, such as social norms, platform rules, social network structures, and the opinion climate, could play a role in influencing their intention to express opinions, particularly in an uncivil way.

In addition, neither positive nor negative emotions are related to exposure to and expression of incivility and polarization. In the observational data, positive and negative emotions are the articulation of verbal expressions that pertain to language or the use of words. The emotions reflect how the opinions are expressed rather than how people feel toward the object after the exposure or expression, which is a key difference in operationalization when using survey and computational analysis to capture individual and collective levels of emotions. The findings suggest that how the uncivil content is emotionally expressed does not affect the extent to which opinions toward the issue diversify, but the uncivil content itself does.

General Discussion

Incivility may escalate nasty talk online, boost aggressiveness, and exacerbate political polarization, all of which are dangerous for healthy democratic development (Anderson et al., 2014). In this study, we utilize two methodological approaches and contextualize incivility in cross-cutting settings to explore whether and how exposure to and expression of online incivility after encountering disagreement online leads to ideological (de)polarization. In the context of the current wave of emigration in Hong Kong, a controversial issue that is related to citizens' partisan leaning and political ideology as well as their various personal and social factors, we revisit the academic debates on the (de)polarizing effects of cross-cutting exposure.

Specifically, we examine ideological polarization as the focal outcome at the individual and collective levels and offer several noteworthy findings. Our results show a very limited deliberative role of incivility in the exposure and expression dimensions. For the exposure dimension of incivility, cross-cutting exposure is related to exposure to uncivil messages (both individual- and collective-level), which is associated with polarization directly (at the collective level) and indirectly (at the individual level, emotional arousal, such as anger or anxiety, plays the mediating role).

For the expression dimension of incivility, at the individual level, cross-cutting exposure is related to expressing uncivil opinions. Individuals may respond to information with which they disagree using hostile language, which further leads to negative emotional arousal and ideological polarization (Kim & Kim, 2019). Interestingly, at the collective level, cross-cutting exposure does not relate to expressing uncivil

opinions, but exposure to uncivil messages does. This finding suggests that individuals may not directly respond to disagreement in an uncivil manner only when the disagreement they encounter is uncivil. Findings from the two studies suggest an aggregated polarizing effect on social media because the more uncivil opinions people express, the more uncivil messages people see. This paints a more worrisome picture of how nasty talk online could polarize public opinion through both *saying* and *seeing*, and when they could reinforce each other. Accordingly, what matters to a healthy dynamic of information diffusion on social media could be moderating cross-cutting exposure/information and reducing the likelihood that people will encounter uncivil content. If cross-cutting exposure can lead to more deliberative information and neutral expression rather than uncivil content, the polarizing effect of cross-cutting exposure through incivility can be mitigated.

One optimistic finding regarding the depolarizing effect of cross-cutting exposure is the mediating role of expressing uncivil opinions at the individual level. Active expression can induce positive emotions and depolarize attitude. This echoes the argument of the expression effect that active expression should be differentiated from passive receiving when examining media effect and information processing (Liang & Ng, 2023). The finding also supports some scholars' argument that censoring uncivil expression harms the open exchange of ideas and deliberative discussion (G. M. Chen, 2017) and that expressive behaviors can heighten individuals' enthusiasm about political affairs (Kosmidis & Theocharis, 2020). However, it is important to note that, similar to exposure to uncivil messages, expressing uncivil opinions also induces negative emotions. The depolarizing effects of the positive emotions that result from expressing uncivil opinions may be counterbalanced by the polarizing effects of the negative emotions that also result, as positive and negative emotions can co-exist (G. M. Chen, 2017; Marcus et al., 2000).

This study also highlights different conceptualizations and operationalizations of emotions in the incivility context. Emotion can be a self-reported feeling, and it can also be a feature of expression. As mentioned above, when it comes to individual-level self-reported data, aroused emotions resulting from exposure to uncivil messages or expressing uncivil opinions play a significant role in mediating the relationship between incivility and polarization. Aroused emotion is a factor that the collective level of social media data did not capture, which may also explain why emotions on discussion forums did not play a role in influencing polarization because emotions are not about how individuals are aroused by the content but how the content is expressed.[6] How people are emotionally aroused by incivility could cause their attitude to be polarized. However, how the uncivil content is emotionally expressed on social media does not affect how the opinion is dispersed.

This study makes several theoretical contributions to the literature. First, it revisits the debate on the relationship between cross-cutting exposure and political polarization by highlighting the mediating role of incivility and specifying polarizing and depolarizing paths. Second, instead of viewing online incivility as a general concept, we differentiate between passive-receiving and active-expressing behaviors. Third, while previous studies on incivility mostly examined the mediating role of negative

emotions, we explore how positive emotions could be generated and play a role in the depolarizing mechanism. Future researchers can continue to utilize different methodological approaches to build on the proposed theoretical model and investigate external- (e.g., political events or issues) or internal- (e.g., motivations) conditional factors that could affect the relationships.

Before concluding the study, several limitations should be acknowledged. First, the causal relationship between cross-cutting exposure and polarization is under debate. Some argue that polarization could reversely affect information exposure (e.g., Stroud, 2010). To address the causality of our model, we draw on a two-wave panel survey to minimize the possibility of a confounding reverse relationship. However, we acknowledge that although the panel survey analysis helps provide evidence for causal inferences, it cannot conclusively confirm the causal direction. Future researchers could include more survey waves or consider experimental designs to further replicate the findings.

Second, we acknowledge the small effect sizes detected. In serial mediation, an indirect effect becomes smaller when more mediators are added. Although the effect sizes of the indirect paths in this study are similar to those reported in prior studies on serial mediation (e.g., Hmielowski & Nisbet, 2016; W. Song & Fox, 2016), it is important to note that these effect sizes are not large. Therefore, caution should be exercised when interpreting the causal-indirect relationships in this study.

Lastly, this study is contextualized in an emerging issue in Hong Kong, which raises the question of generalizing the current findings to other politically polarized issues. As mentioned earlier, Hong Kong's emigration issue is different from longstanding issues like immigration or abortion in the U.S., which have exhibited clear divides over decades and strong alignment with partisan and ideological affiliations. Therefore, it is important to examine the proposed relationships across different types of issues and political contexts using both individual and collective data. We posit that our model and proposed relationships are amenable to being tested in other polarized issues, with their effects potentially transferrable to those issues given that the hypotheses we propose are based on literature that investigates relationships within clearly divided societies, primarily in Western contexts. Thus, we suggest that in cases where political issues are similarly or more polarized and attitudes and party stances converge, the patterns between variables such as exposure, expression, emotional arousal, and polarization may exhibit even greater significance. This warrants further research to deepen our understanding of these dynamics.

In sum, this study revisits the debate on whether cross-cutting exposure leads to deliberation or backfires by incorporating the role of online incivility, a growing phenomenon in today's networked public sphere. Although there are some potential benefits to allowing uncivil expression in terms of emotional engagement and political enthusiasm, in the context of cross-cutting exposure, the overall deliberative role of incivility on social media appears to be limited due to its polarizing effects. We need to be cautious about the detrimental impact of incivility on social cohesion and deliberative democracy when people encounter disagreement on social media.

Declaration of Conflicting Interests

Funding

Open Practice

The survey data that support the findings of Study 1 are openly available. https://osf.io/nrp9w/?view_only=8cc96143448f49f7bf9292528b2d9d29. The online forum data that support the findings of Study 2 will be shared on reasonable request to the corresponding author due to the sensitivity of the issue topic. Data from online forums are public and can contain details that pose privacy concerns.

ORCID iDs

Hsuan-Ting Chen https://orcid.org/0000-0003-3140-5169

Yunya Song https://orcid.org/0000-0001-5159-1689

Jing Guo https://orcid.org/0000-0001-8707-1124

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Unlike attitudes toward immigration or abortion in the US, Hong Kongers' attitudes toward the recent wave of emigration do not completely align with their partisan (pro-establishment and pro-democracy) or ideological (conservative vs. liberal) affiliations. We leveraged this unique feature of the Hong Kong emigration issue to examine ideological polarization with two methodological approaches. Such efforts could help to explain the (de)polarizing effects of cross-cutting exposure in an emerging and controversial issue-specific context.
2. In the Hong Kong context, we operationalize "social media platform" as an umbrella term that includes (1) social network sites, such as Facebook and X (Twitter); (2) discussion forums, such as LIHKG and The Hong Kong Discuss Forum; and (3) instant messaging apps, such as WhatsApp and WeChat.
3. In the U.S., ideological polarization is captured by the distance between the two opposite sides of political ideologies or issue opinions. Ideological polarization emerges on a range of political issues in the US context, such as abortion rights, gun control, immigration policy, and COVID-19 pandemic control (Hart et al., 2020).
4. We include information about the indirect paths in the saturated model in Appendix E.
5. Three discrete negative emotions (i.e., anger, anxiety, and sadness) were also counted,

but they are not analyzed separately for this study as we focus on positive and negative emotions.

6. Computational scientists typically conduct sentiment analysis with various natural language processing techniques and dictionaries to capture the emotions expressed. One may argue that positive and negative emotions (i.e., sentiment) captured in the text can suggest affective polarization (Yarchi et al., 2021); however, this approach does not take into account whether that sentiment is expressed toward any particular group (Overgaard et al., 2024), which is the main idea of affective polarization (i.e., negative feelings toward outparty members; Iyengar et al., 2019).

References

Anderson, A. A., Brossard, D., Scheufele, D. A., Xenos, M. A., & Ladwig, P. (2014). The "nasty effect:" Online incivility and risk perceptions of emerging technologies. *Journal of Computer-Mediated Communication*, *19*(3), 373–387. https://doi.org/10.1111/jcc4.12009

Aynaud, T., Fleury, E., Guillaume, J. L., & Wang, Q. (2013). Communities in evolving networks: definitions, detection, and analysis techniques. In Dynamics on and of complex networks, volume 2: applications to time-varying dynamical systems (pp. 159–200). Birkhäuser. https://doi.org/10.1007/978-1-4614-6729-8_9

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of National Academy of Sciences of the United States of America*, *115*(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barnidge, M. (2017). Exposure to political disagreement in social media versus face-to-face and anonymous online settings. *Political Communication*, *34*(2), 302–321. https://doi.org/10.1080/10584609.2016.1235639

Barnidge, M. (2020). Testing the inadvertency hypothesis: Incidental news exposure and political disagreement across media platforms. *Journalism*, *21*(8), 1099–1118. https://doi.org/10.1177/1464884920915373

Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de)polarization: Reinforcing spirals in the 2016 US election. *Information, Communication & Society*, *21*(7), 940–958. https://doi.org/10.1080/1369118X.2018.1444783

Berry, J. M., & Sobieraj, S. (2014). *The outrage industry: Political opinion media and the new incivility*. Oxford University Press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *2*, 993–1022. https://doi.org/10.7766/orbit.v1.2.44

Brooks, D. J., & Geer, J. G. (2007). Beyond negativity: The effects of incivility on the electorate. *American Journal of Political Science*, *51*(1), 1–16. https://doi.org/10.1111/j.1540-5907.2007.00233.x

Chan, M., Chen, H.-T., & Lee, F. L. F. (2017). Examining the roles of mobile and social media in political participation: A cross-national analysis of three Asian societies using a communication mediation approach. *New Media & Society*, *19*(12), 2003–2021. https://doi.org/10.1177/1461444816653190

Chan, M., Chen, H.-T., & Lee, F. L. F. (2021). Cross-cutting discussion on social media and online political participation: A cross-national examination of information seeking and social accountability explanations. *Social Media & Society*, *7*(3), 1–13. https://doi.org/10.1177/20563051211035697

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

Chen, G. M., & 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. https://doi.org/10.1080/08838151.2016.1273922

Chen, H.-T., Kim, Y., & Chan, M. (2022). Just a glance, or more? Pathways from counter-attitudinal incidental exposure to attitude (de)polarization through response behaviors and cognitive elaboration. *Journal of Communication*, *72*(1), 83–110. https://doi.org/10.1093/joc/jqab046

Chen, H.-T., & Lin, J.-S. (2021) Cross-cutting and like-minded discussion on social media: The moderating role of issue importance in the (de)mobilizing effect of political discussion on political participation. *Journal of Broadcasting and Electronic Media*, *65*(1), 135–156. https://doi.org/10.1080/08838151.2021.1897822

Chen, W., Pacheco, D., Yang, K. C., & Menczer, F. (2021). Neutral bots probe political bias on social media. *Nature Communication*, *12*(1), Article 5580. https://doi.org/10.1038/s41467-021-25738-6

Dalton, R. J. (1987). Generational change in elite political beliefs: The growth of ideological polarization. *The Journal of Politics*, *49*(4), 976–997. https://doi.org/10.2307/2130780

De Francisci Morales, G., Monti, C., & Starnini, M. (2021). No echo in the chambers of political interactions on Reddit. *Scientific Reports*, *11*, Article 2818. https://doi.org/10.1038/s41598-021-81531-x

Delli Carpini, M. X., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annual Review of Political Science*, *7*(1), 315–344. https://doi.org/10.1146/annurev.polisci.7.121003.091630

Dillard, J. P., & Shen, L. (2007). Self-report measures of discrete emotions. In R. A. Reynolds, R. Woods & J. D. Baker (Eds.), *Handbook of research on electronic surveys and measurement* (pp. 330–333). Idea Group Inc.

DiMaggio, P., Evans, J., & Bryson, B. (1996). Have American's social attitudes become more polarized? *American Journal of Sociology*, *102*(3), 690–755. https://doi.org/10.1086/230995

Festinger, L. A. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Fredrickson, B. L., & Branigan, C. (2001). Positive emotions. In T. J. Mayne & G. A. Bonnano (Eds.), *Emotion: Current issues and future directions* (pp. 123–151). Guilford Press.

Gervais, B. T. (2014). Following the news? Reception of uncivil partisan media and the use of incivility in political expression. *Political Communication*, *31*(4), 564–583. https://doi.org/10.1080/10584609.2013.852640

Gil, de, Zúñiga, H., Molyneux, L., & Zheng, P. (2014). Social media, political expression, and political participation: Panel analysis of lagged and concurrent relationships. *Journal of Communication*, *64*(4), 612–634. https://doi.org/10.1111/jcom.12103

Guo, J., & Chen, H.-T. (2022). How Does Multi-Platform Social Media Use Lead to Biased News Engagement? Examining the Role of Counter-Attitudinal Incidental Exposure, Cognitive Elaboration, and Network Homogeneity. *Social Media + Society*, *8*(4). https://doi.org/10.1177/20563051221129140

Gutmann, A., & Thompson, D. (1998). *Democracy and disagreement*. Harvard University Press.

Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science communication*, *42*(5), 679-697. https://doi.org/10.1177/1075547020950735

Hmielowski, J. D., Hutchens, M. J., & Cicchirillo, V. J. (2014). Living in an age of online incivility: Examining the conditional indirect effects of online discussion on political flaming. *Information, Communication & Society*, *17*(10), 1196–1211. https://doi.org/10.1080/1369118X.2014.899609

Hmielowski, J. D., & Nisbet, E. C. (2016). "Maybe yes, maybe no?": Testing the indirect relationship of news use through ambivalence and strength of policy position on public engagement with climate change. *Mass Communication and Society*, *19*(5), 650–670. https://doi.org/10.1080/15205436.2016.1183029

Hopp, T., & Vargo, C. J. (2017). Does negative campaign advertising stimulate uncivil communication on social media? Measuring audience response using big data. *Computers in Human Behavior*, *68*, 368–377. https://doi.org/10.1016/j.chb.2016.11.034

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Isen, A. (1990). The influence of positive and negative affect on cognitive organization: Some implications for development. In N. Stein, B. Leventhal & T. Trabasso (Eds.), *Psychological and biological approaches to emotion* (pp. 75–94). Erlbaum.

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*, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034

Kim, Y., & Kim, Y. (2019). Incivility on Facebook and political polarization: The mediating role of seeking further comments and negative emotion. *Computers in Human Behavior*, *99*, 219–227. https://doi.org/10.1016/j.chb.2019.05.022

Kosmidis, S., & Theocharis, Y. (2020). Can social media incivility induce enthusiasm? Evidence from survey experiments. *Public Opinion Quarterly*, *84*(Suppl. 1), 284–308. https://doi.org/10.1093/poq/nfaa014

Kubin, E., & Sikorski, C. v. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, *45*(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Lee, F. L. F., Liang, H., & Tang, G. K. Y. (2019). Online incivility, cyberbalkanization, and the dynamics of opinion polarization during and after a mass protest event. *International Journal of Communication*, *13*, 4940–4959.

Lee, H., & Kwak, N. (2014). The affect effect of political satire: Sarcastic humor, negative emotions, and political participation. *Mass Communication and Society*, *17*(3), 307–328. https://doi.org/10.1080/15205436.2014.891133

Liang, H., & Ng, Y. M. M. (2023). The expression effects of uncivil disagreement: The mechanisms of cognitive dissonance and self-perception. *Human Communication Research*, *49*(3), 483–507. https://doi.org/10.1093/hcr/hqac032

Liang, H., & Zhang, X. (2021). Partisan bias of perceived incivility and its political consequences: Evidence from survey experiments in Hong Kong. *Journal of Communication*, *71*(3), 357–379. https://doi.org/10.1093/joc/jqab008

Lin, H., Wang, Y., Lee, J., & Kim, Y. (2023). The effects of disagreement and unfriending on political polarization: A moderated-mediation model of cross-cutting discussion on affective polarization via unfriending contingent upon exposure to incivility. *Journal of Computer-Mediated Communication*, *28*(4), Article zmad022. https://doi.org/10.1093/jcmc/zmad022

Marcus, G. E., Neuman, W. R., & MacKuen, M. B. (2000). *Affective intelligence and political judgment*. University of Chicago Press.

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, *59*(4), 128–145. https://doi.org/10.1111/ajps.12089

McLaughlin, B., Holland, D., Thompson, B. A., & Koenig, A. (2020). Emotions and affective polarization: How enthusiasm and anxiety about presidential candidates affect interparty attitudes. *American Politics Research*, *48*(2), 308–316. https://doi.org/10.1177/1532673X19891423

Muddiman, A. (2017). Personal and public levels of political incivility. *International Journal of Communication*, *11*, 3182–3202.

Mutz, D. C. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, *96*(1), 111–126. https://doi.org/10.1017/S0003055402004264

Mutz, D. C. (2015). *In-your-face politics: The consequences of uncivil media*. Princeton University Press.

Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), Article 026113. https://doi.org/10.1103/PhysRevE.69.026113

Newman, M. E., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, *64*(2), Article 026118 https://doi.org/10.1103/PhysRevE.64.026118

Omdahl, B. L. (2014). *Cognitive appraisal, emotion, and empathy*. Psychology Press.

Overgaard, C. S. B., Lukito, J., & Soorholtz, K. (2024, May). The Manifestation of Affective Polarization on Social Media: A Cross-Platform Supervised Machine Learning Approach. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 18, pp. 1160–1173). https://doi.org/10.1609/icwsm.v18i1.31380

Oz, M., Zheng, P., & Chen, G. M. (2018). Twitter versus Facebook: Comparing incivility, impoliteness, and deliberative attributes. *New Media & Society*, *20*(9), 3400–3419. https://doi.org/10.1177/1461444817749516

Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, *4*(1), 9-27. https://doi.org/10.1177/14614440222226244

Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, *6*(2), 259–283. https://doi.org/10.1177/1461444804041444

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied Psychology*, *41*(4), 359–376. https://doi.org/10.1111/j.1464-0597.1992.tb00712.x

Phillips, T., & Smith, P. (2004). Emotional and behavioural responses to everyday incivility: Challenging the fear/avoidance paradigm. *Journal of Sociology*, *40*(4), 378–399. https://doi.org/10.1177/1440783304048382

Pingree, R. J. (2007). How messages affect their senders: A more general model of message effects and implications for deliberation. *Communication Theory*, *17*(4), 439–461. https://doi.org/10.1111/j.1468-2885.2007.00306.x

Rains, S. A., Kenski, K., Coe, K., & Harwood, J. (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. https://doi.org/10.1111/jcc4.12191

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Schudson, M. (1997). Cultural memory: A special issue. *The Communication Review*, *2*(1), 1–2. https://doi.org/10.1080/10714429709368546

Song, W., & Fox, J. (2016). Playing for love in a romantic video game: Avatar identification, parasocial relationships, and Chinese women's romantic beliefs. *Mass Communication and Society*, *19*(2), 197–215. https://doi.org/10.1080/15205436.2015.1077972

Song, Y., Lin, Q., Kwon, H., Choy, C. H. Y., & Xu, R. (2022). Contagion of offensive speech online: An interactional analysis of political swearing. *Computers in Human Behavior*, *127*, Article 107046. https://doi.org/10.1016/j.chb.2021.107046

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, *60*(3), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

Stroud, N. J. (2017). Helping newsrooms work toward their democratic and business objectives. In P. J. Boczkowski & C. W. Anderson (Eds.), *Remaking the news: Essays on the future of journalism scholarship in the digital age* (pp. 157–176). MIT Press.

Sun, Q., Wojcieszak, M., & Davidson, S. (2021). Over-time trends in incivility on social media: Evidence from political, non-political, and mixed sub-Reddits over eleven years. *Frontiers in Political Science*, *3*, Article 741605. https://doi.org/10.3389/fpos.2021.741605

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A bad workman blames his tweets: The consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of Communication*, *66*(6), 1007–1031. https://doi.org/10.1111/jcom.12259

Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., & Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *Journal of Politics*, *73*(1), 156–170. https://doi.org/10.1017/S0022381610000939

Wang, J., & Sundar, S. S. (2022). Liking versus commenting on online news: Effects of expression affordances on political attitudes. *Journal of Computer-Mediated Communication*, *27*(6), Article zmac018. https://doi.org/10.1093/jcmc/zmac018

Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, *81*, 73–83. https://doi.org/10.1016/j.chb.2017.11.041

Wojcieszak, M., & Garrett, R. K. (2018). Social identity, selective exposure, and affective polarization: How priming national identity shapes attitudes toward immigrants via news selection. *Human Communication Research*, *44*(3), 247–273. https://doi.org/10.1093/hcr/hqx010

Wojcieszak, M., & Price, V. (2009). What underlies the false consensus effect? How personal opinion and disagreement affect perception of public opinion. *International Journal of Public Opinion Research*, *21*(1), 25–46. https://doi.org/10.1093/ijpor/edp001

Wojcieszak, M., & Price, V. (2010). Bridging the divide or intensifying the conflict? How disagreement affects strong predilections about sexual minorities. *Political Psychology*, *31*(3), 315–339. https://doi.org/10.1111/j.1467-9221.2009.00753.x

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. https://doi.org/10.1080/10584609.2020.1785067

Zhao, N., Jiao, D., Bai, S., & Zhu, T. (2016). Evaluating the validity of simplified Chinese version of LIWC in detecting psychological expressions in short texts on social network services. *PLOS ONE*, *11*(6), Article e0157947. https://doi.org/10.1371/journal.pone.0157947

Author Biographies

Hsuan-Ting Chen, Ph.D., is an Associate Professor at the School of Journalism and Communication, The Chinese University of Hong Kong. Her research centers on the utilization of emerging media and its impact on deliberation and political communication processes.

Yunya Song, Ph.D., is a Professor at Hong Kong Baptist University. Her research cuts across global communication, digital media, computer-mediated networks, social media analytics, cyber-psychology and behavior.

Jing Guo, Ph.D., is an Assistant Professor at Hong Kong Shue Yan University. Her research focuses on political communication and new media.