



Contagion of offensive speech online: An interactional analysis of political swearing

Yunya Song^{a,*}, Qinyun Lin^b, K. Hazel Kwon^{c,2}, Christine H.Y. Choy^{d,3}, Ran Xu^e

^a Department of Journalism, Hong Kong Baptist University, Hong Kong

^b Center for Spatial Data Science, University of Chicago, United States

^c Walter Cronkite School of Journalism and Mass Communication, Arizona State University, United States

^d Department of Art and Design, The Hong Kong Baptist University, Hong Kong

^e Department of Allied Health Sciences, University of Connecticut, United States

ARTICLE INFO

Keywords:

Swearing
Reciprocity
Mimicry
Online incivility
Contagion
Anti-social

ABSTRACT

Despite their positive effects in promoting participatory politics, digital publics have also manifested an offensive vernacular culture. This study takes a social network analytic approach to explain the contagion of offensive speech in online discussion contexts. The study examines four social interactional mechanisms underlying a user's adoption of political swearing: generalized reciprocity, direct reciprocity, leader-mimicry, and peer-mimicry. The empirical context of this study is a highly popular online discussion forum in Hong Kong. The study examines the effects of social interactional mechanisms on the occurrences of political swearing by analyzing five years of user comments. Findings show that peer-mimicry contributes to the contagion process the most, followed by generalized reciprocity and direct mimicry. The study demonstrates how individual-level speech behaviors spiral into a collective norm that potentially hinders a healthy discussion culture in mediated social spaces.

The discussion culture of a digital platform is constructed by how users attune their conversational styles to one another and meld them into a shared communicative practice (Toma, 2014). Unfortunately, recent digital social spaces have often shown a distorted convergence of conversational styles by exhibiting overly intensive emotional exchanges, online harassment, hate speech, and verbal revenge (Kwon & Gruzd, 2017; Song et al., 2021). The web facilitates the prevalence of Internet trolls (March, 2019), cyberbullying (Fearn, 2017), hate speech (Ortiz, 2019), and more covert forms of offensive speech (e.g., sarcastic, ironic comments) that are appropriated to hurt individuals subtly (Frenda, 2018). When it comes to political conversations, toxic comments and antithetical expressions have become prevalent in many digital platforms, often tainting the democratic potential of digital

participatory culture (Coe et al., 2014).

This study explores how individual-level offensive utterances can collectively spiral into a toxic vernacular culture online by making a case for political swearing as an example of the contagion of offensive speech. While not all swearing is necessarily offensive, anonymous political swearing in the mediated environment is frequently paired with flaming, trolling, cyberbullying, and retaliation. Existing studies have suggested that online users may engage in swearing on the spur of the moment (Ng et al., 2020; Song et al., 2021), but to what extent is such spontaneity a product of social contagion? Also, to what extent is an act of swearing attributed to factors beyond intra-individual tendencies? If situational factors matter beyond individual traits, what are the social interactional conditions that contribute to the spread of swearing

* Correspondence author. Department of Journalism, Hong Kong Baptist University, Communication and Visual Arts Building, 5 Hereford Road, Kowloon, Hong Kong.

E-mail address: yunyasong@hkbu.edu.hk (Y. Song).

¹ Her research on digital media and political communication has appeared in, among other journals, *New Media & Society*, *International Journal of Press/Politics*, and *Social Science Computer Review*. She is the coeditor of *The Evolving Landscape of Media and Communication in Hong Kong* (City University of Hong Kong Press, 2018).

² Her publications have appeared in multiple journals, including the *Journal of Computer-Mediated Communication*, *CyberPsychology, Behavior & Social Networking*, *Computers in Human Behaviors*, *American Behavioral Scientists*, and *Social Science Computer Review*.

³ Her research has been published in multiple journals, including the *Mobile Media & Communication*, *Telecommunications Policy*, *Public Relations Review*, *International Journal of Conflict Management*, and *Journalism & Mass Communication Quarterly*.

behaviors?

To empirically examine social contagion mechanisms of offensive online speech, we employ a network analytic approach to large-scale discussion posts and to replying comments collected between 2014 and 2018 from Hong Kong Golden Forum (HKGolden), one of the most popular online forums in Hong Kong during that time period. In Hong Kong, the time period was noteworthy due to the rise of the Umbrella Movement in 2014, one of the largest political movements in Hong Kong history, and its continuing ramifications. The outbreak of the Umbrella Movement in 2014, the 79-day occupation of key sites in the city demanding more transparent elections, popularized the principle of civil disobedience as a struggle for justice, and marked a major turning point in the direction of Hong Kong's political culture. Newly emerged localism groups criticized the traditional approach of striving for democracy (known by the slogan "peace, rational, non-violence and non-profanity") as too conciliatory (Su, 2019). Five years on, Hong Kong was again in the throes of mass protests that turned out to involve violent confrontations between protesters and police (Cheung, 2019).

Recognizing the importance of the Umbrella Movement in global social movement history, this study takes a slightly different angle by focusing on how political discussion culture has revealed itself in the aftermath of the movement. Anecdotal cases have suggested Hong Kong digital spaces have become notably hostile in the post-Umbrella Movement (post-UM) period (F. Lee, 2018). This study contends that, while the political movement has revealed Hong Kong's sociopolitical conflict and heightened tension with Mainland China, networked speech culture eventually evolves through the ways in which users interact. That is, a rigorous understanding of social interactional mechanisms may offer more fundamental insights on the emergence of offensive speech culture than solely attributing it to political causes. While the empirical context is Hong Kong, the study advances generalizable knowledge on how offensive speech becomes an online community's norm by introducing and testing a theory-driven framework of the contagion mechanisms.

1. Research background

1.1. Political swearing online as individual speech and a collective norm

Antagonistic speech has become increasingly prevalent online despite most digital platforms maintaining policies for civility, inclusivity, and respect (Kwon & Gruzd, 2017). Political swearing has become part of reactionary web vernaculars (Coe et al., 2014). Broadly speaking, swearing refers to utterances of "offensive or aggressive emotional languages that are inhibited by social convention or aversion" (Jay, 2009, p. 153). An expression of hostility often occurs when a focal actor encounters uncongenial information that challenges his or her existing belief and thus feels an immediate urge to defend oneself (De Dreu & van Knippenberg, 2005). Swearing is an easy means to let emotions out, especially in an anonymous textual discussion setting (Herring et al., 2002; Lampe et al., 2014). Classical anti-social behavior theories centered on anonymity, deindividuation, and disinhibition have explained motivations for offensive online speech (e.g., Suler, 2004). A notable difference between offensive online speech and anti-social offline behavior is that the former is less likely to be premeditated. Most recently, Tuters and Hagen (2020) have demonstrated that anonymous online communities such as 4chan not only breed antagonistic speech but also turn it into a memetic culture.

In addition to anonymity and emotional urge, social contagion theory can explain how individuals collectively create an offensive culture. Social contagion refers to a social transmission process where individuals act as a stimulus to spread ideas or behavioral patterns in the wider community (Marsden, 1998). It is a process wherein a collective of individuals adopts ideas, affects, and behaviors that have been communicated through social interactions (Bastiampillai et al., 2013). Empirically, social contagion has been tested by examining whether one's behavior (or emotion) converges with others' behaviors (or

emotions). Individuals may engage in a relationship that satisfies other social needs, while this relationship may inadvertently expose the focal individual to a certain behavior, leading to his or her adoption of the behavior. Furthermore, social proximity has been a defining factor that leads to behavioral contagion: the stronger the friendship, the greater the likelihood of the peer's behavior becoming contagious (Larson et al., 2007). Importantly, social dynamics in peer networks are multilayered, embedded in not only a dyadic relationship but also a clique (or a tightly-knit group of friends), a broader cohort, and even a crowd (Brown & Klute, 2003).

Meanwhile, emergent norm theory explains how social contagion breeds new expectations and norms in crowds. Likemindedness, anonymity, and shared emotion interplay to shape crowd behaviors (Arthur, 2013; Turner & Killian, 1987). An elaborated social identity model of the crowd (Reicher, 1996) has further suggested that norms become contagious when there is a demand to perform a collective action to show group members' shared values, goals, and identities to outsiders (Reicher & Stott, 2011).

A large-scale analysis of moderated comments on Reddit identified different norms across disparate online communities that are reflective of what Reddit values and how widely held those values are (Chandrasekharan et al., 2018). Research has also shown that social norms that allow perpetrators to repeatedly attack victims are contagious in online social networks (e.g., Ballard & Welch, 2017; Wong et al., 2018). Using a combination of questionnaire and machine learning methods, Yokotani and Takano (2021) estimated the status of cyberbullying perpetrators and victims in a Japanese online chat platform, suggesting social norms that condone cyberbullying were contagious via online peer networks. Hmielowski et al. (2014) scrutinized the way computer-mediated communication socializes people into seeing verbal aggression as acceptable, and even radicalizes them into adopting similar behavior at times. Shmargad et al.'s (2021) analysis of large-scale news comments showed that online commenters are responsive to the behaviors of other discussion community members, which influence the likelihood of repeated incivility.

1.2. What drives offensive speech online

Existing literature has suggested a set of individual motivations for engaging in anti-social behaviors, which include alleviating boredom (Varjas et al., 2010), fun-seeking (Shachaf & Hara, 2010), and venting (S. Lee & Kim, 2015). Research has extended motivational inquiries to various forms of online communication. Notably, this line of research focuses on "flaming" in political discussion contexts, defined as hostile, insulting expressions (Cho & Kwon, 2015) which may "inflict harm to a person or an organization resulting from uninhibited behavior" (Alonso & Aiken, 2004, p. 205). Overall, research on flaming has probed how individual and situational factors contribute to political flaming.

Regarding individual factors, S. Lee and Kim (2015) interviewed social media users, revealing that malicious commenting behaviors were in large part caused by a desire to express anger or by feelings of inferiority. Rozin and Royzman (2001) reviewed the evidence for "negativity bias", suggesting that negative entities are more contagious than positive affect due to human beings' innate predispositions and experience. Studies also suggested that users with a high tendency of verbal aggression—"the tendency to attack another person's self-concept rather than the topic of the conversation"—showed a stronger intention of political flaming (Hmielowski et al., 2014, p. 1201). Political behavioral factors such as participation, offline political discussion network, and exposure to profane media content were also associated with online political flaming (Coyne et al., 2011; Hutchens et al., 2015).

Regarding situational factors, the literature has consistently shown anonymity to be an important situational factor underlying offensive political interactions. For example, Rowe (2014) showed that anonymous or pseudonymous news comment sites contained double the amount of political swearing as Facebook pages where commenters' real

identities were traceable. Similarly, [Cho and Kwon \(2015\)](#) analyzed a large-scale dataset of online comments, and found that users engaged in less political swearing in real-name-based social platforms than in platforms that allowed pseudonym-based social interactions. In addition to anonymity, a variety of discursive contextual factors have contributed to political flaming, such as the subject of discussion ([Kwon & Cho, 2017](#); [Rossini et al., 2021](#)) and a perceived threat or provocation during discussion ([Hutchens et al., 2015](#)). J. [Cheng et al.'s \(2017\)](#) large-scale analysis of trolling in news comments echoed the role of socialization in cultivating flaming behaviors: discussion contexts that presented less thoughtful posts would breed less thoughtful responses. These studies reaffirmed the role of norms in shaping the discussion styles of the community. The creation of a “flaming norm” ([Moor et al., 2010, p. 1538](#)) resonates with a “broken windows hypothesis,” which postulates that unmoderated anti-social behaviors may lead to the breakdown of a whole community ([Wilson & Kelling, 1982](#)).

To summarize, there is a rich body of scholarship on both individual and situational factors associated with offensive online speech. What has been missing in the current literature, however, are the underlying social interactional processes, i.e., the ways in which individuals over time shift their speech styles to behave like others. To understand how offensive speech is transmitted user-to-user and eventually forms a collective community norm, we set out to expand on the social contagion mechanisms of anti-social behavior.

2. Theoretical framework: social contagion mechanisms of offensive online speech

Social contagion has become an established social science concept, especially in the field of marketing and diffusion research. Traditional diffusion models, for example, have estimated parameters that separate the broadcasting mechanism from the word-of-mouth effect on product adoption ([Bass, 2004](#); [Rogers, 2003](#)). While researchers have developed various social contagion models to elaborate the network mechanisms underlying individuals' attitudinal or behavioral adoptions, there are few proposed mechanisms that specifically tackle the contagion of anti-social behaviors. Probably most widely known is [Tsvetkova and Macy's \(2015\)](#) anti-social contagion framework. Tsvetkova and Macy's model assumes that an anti-social contagion process involves three types of actors –instigators (who initiate a bad behavior), targets (who are offended by instigators), and observers (who witness someone else engaging in a bad behavior) –whose relational dynamics constitute two mechanisms of anti-social contagion: generalized reciprocity and third-party influence ([Tsvetkova & Macy, 2015](#)). Expanding on [Tsvetkova and Macy's \(2015\)](#) anti-social contagion framework, we offer four contagion mechanisms of offensive speech online: generalized

reciprocity, direct reciprocity, leader-mimicry, and peer-mimicry (see [Fig. 1](#)).

2.1. Generalized reciprocity

The generalized reciprocity effect describes “a pay-it-forward cascade of contagious malevolence” to retaliate “not against the perpetrator but against innocent others (p. 38).” ([Tsvetkova & Macy, 2015, pp. 38–40](#)). In an online discussion setting, perception of interactional injustice ([Bies & Moag, 1986](#)) can trigger generalized reciprocity of offensive speech. Online communities have discursive norms and rules based on a shared understanding of acceptable ways of interacting ([Yee et al., 2007](#)). A study of over 6000 online newspaper comments found that the affirmation of initial incivility by descriptive and injunctive norms engenders repeated incivility in an online discussion, leading to the conclusion that online incivility is a “dynamic, normative process” that is responsive to “proximate incivility” ([Shmargad et al., 2021, p. 1](#)).

Experiencing a breach of shared interactional norms can engender the observer's negative affect toward the community, as if the interactional injustice is characteristic of the entire community ([Skarlicki & Folger, 1997](#)). Previous studies have examined the underlying affective drivers of paying behavior forward. While positive emotions drive paying-it-forward altruistic behavior, receiving such behavior as greedy likely arouses negative effect ([Sanfey et al., 2003](#)), which in turn typically has larger influence than positive affect. Negative perceptions then motivate the targeted user to be less attentive to the consequences of verbal attacks when there are alternatives ([Tedeschi & Felson, 1994](#)). One easy way to release the negative effect would be to replicate an unfair act, which, in turn, might lead to a similar perception and reaction by the third user, thus potentially resulting in a chain of interactional injustice ([Patterson, 1982](#)). That is, an incivility spiral may begin when the targeted user enacts revenge by returning similar verbal aggression toward another “innocent” user. Generalized reciprocity of offensive speech may not necessarily intend to harm the target but merely to release a negative effect aroused by the verbal attack they received ([Kim & Tsvetkova, 2020](#)). Therefore, the mechanism of generalized reciprocity characterizes the pay-it-forward behavior of a target (i.e., the offended) that intends to release the negative effect. Based on the norms of reciprocity, when people perceive that the behavior is prevalent among the community members, they are more likely to conform and pay it forward through intensive interaction.

Hypothesis 1 (Generalized Reciprocity). Being targeted by an offensive comment increases the likelihood of the targeted user subsequently forwarding an offensive comment to a third user.

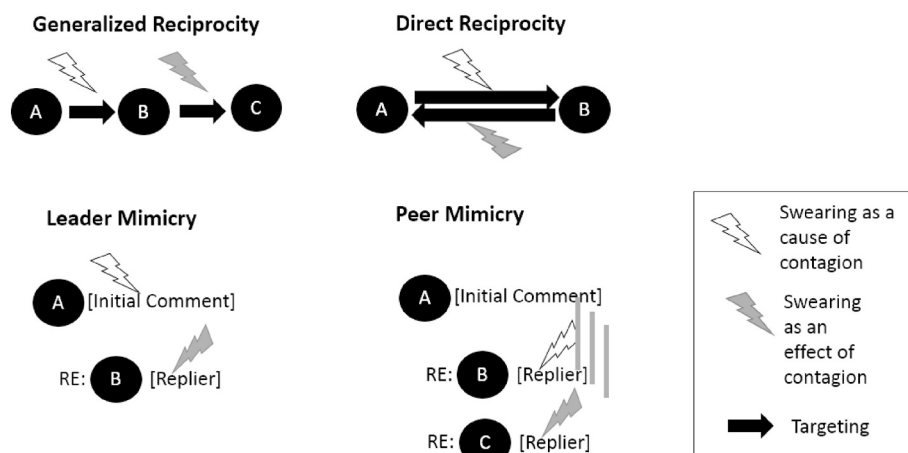


Fig. 1. Contagion mechanisms of offensive speech online.

2.2. Direct reciprocity

While generalized reciprocity pertains to the target's desire to pay it forward, direct reciprocity means that the target responds directly to the instigator. [Tsvetkova and Macy \(2015\)](#) intentionally dismiss the process of direct reciprocity in their framework because they assume that direct reciprocity occurs in a closed network, whereas social contagion necessitates an open network where behaviors diffuse to a wider population.

The contagion of offensive speech in an online discussion, however, entails both generalized and direct social interactions. It is because, an interaction between the instigator and the target is not confined in a closed dyad but embedded in an open discussion context where a third person may easily jump into the conversation to evolve the dyad into a triadic or larger conversational network.

In fact, a direct reaction to the stimulus has been the baseline premise of many social contagion models. For example, the independent cascade model (ICM), one of the most widely studied contagion models, formulates a monotonic effect of direct exposure to an "infected" friend on one's own chance of being infected by the information, attitude, or idea ([Hodas & Lerman, 2014](#); [Kempe et al., 2003](#)). Other models have similarly focused on the direct exposure effect yet assume a non-monotonic effect of exposure. For example, linear threshold model ([Granovetter, 1978](#); [Valente, 1996](#)) suggests that an individual adopts an idea or behavior when the proportion of friends adopting it passes one's personal threshold of resistance. [Kwon et al. \(2014\)](#) expanded the personal network threshold model in the context of Facebook group formation, finding that the frequency of friends' direct messaging that promoted an advocacy group increased a user's likelihood of joining the group. Meanwhile, other studies in social media environments have shown that the repeated exposure effect has a carrying capacity in that the repeated exposure had an inverted curvilinear effect on contagion ([Romero et al., 2011](#); [Ver Steeg et al., 2011](#)).

These studies have commonly suggested that an individual's adoption of an idea or behavior is the product of contagion through direct interactions with the "infected" others. In political flaming contexts, the literature has similarly alluded to the importance of direct interaction effect. For example, "reactive aggression" has been discussed as a situational factor that engenders offensive speech online: Users post a hostile message as a response to a perceived provocation, especially when they receive "comments that challenge their beliefs directed at them personally ... using a person's name (the user's screen name or real name) in the comment" ([Hutchens et al., 2015](#), p. 1205).

Based on the social contagion models that focus on the direct contact effect and the reactive aggression effect found in the political flaming context, we consider direct reciprocity to be another important mechanism for the contagion of offensive speech online.

Hypothesis 2 (Direct Reciprocity). Being targeted by an offensive comment increases the likelihood of the targeted user offensively responding to the instigator.

2.3. Verbal mimicry

While reciprocity effects are centered on the targeted actor's reaction to the attack, [Tsvetkova and Macy \(2015\)](#) define "third-party influence" as a mechanism of imitation by untargeted actors. The theoretical root of third-party influence mechanisms is social learning theory, which posits that observational learning makes people adopt the observed behavior ([Bandura, 1978](#)). [Seering et al. \(2017\)](#), for example, attribute digital incivility to observational learning more than to anonymity or an individual mood state. That is, the prevalence of offensive speech can make users perceive it to be a "descriptive norm" of the community, creating a mental shortcut that "if everyone is doing it, it must be a sensible thing to do" ([Cialdini et al., 1991](#), p. 1015). [Kwon et al. \(2014\)](#)'s social contagion study on Facebook has incorporated the third-party influence

perspective by conducting a field experiment on participants who receive Facebook's automated "social message" (p. 1349) about who and how many of their friends joined a promoted campaign. The results suggested that indirect exposure rate was positively associated with one's likelihood of joining the campaign.

When it comes to the contagion of pro-social behaviors, [Tsvetkova and Macy \(2015\)](#) suggest that third-party influence should diminish over time because people tend to believe that help is no longer needed once the pro-social behavior prevails. Interestingly, however, this diminishing effect is not applicable to anti-social behavior. In contrast, the rampancy of anti-social behavior encourages observers to adopt the behavior, given the perception that the behavior has become common and thus acceptable ([Tsvetkova & Macy, 2015](#)).

Developed from the logic of third-party influence and social learning theory, we propose verbal mimicry as a mechanism of offensive speech contagion online. Verbal mimicry reflects convergence in speech style—or writing in the context of text-based community—by synchronizing syntax, speech rate and utterance ([Chartrand & van Baaren, 2009](#)). [Kwon and Gruzd \(2017\)](#) offered evidence that anonymous YouTube users mimicked others' political swearing. Based on the two types of swearing, interpersonal and untargeted swearing, they (2017) found that a lead post and a subsequent replying comment had disproportionate effects on engendering swearing mimicry. Specifically, untargeted swearing by a lead post prompted both targeted and untargeted swearing in its replies, while a replying comment influenced the subsequent peer comments only for the same type of swearing—either targeted or untargeted. Expanding on [Kwon and Gruzd \(2017\)](#), this study distinguishes two types of mimicry effect, by defining the imitation of a lead post as *leader mimicry* (given that the lead post initiates the discussion thread followed by subsequent comments) and the imitation of other replying comments as *peer mimicry* (given that these comments together constitute an interactive discussion network that not only responds to the lead post but also exchanges opinions among themselves). Two hypotheses are thus posited:

Hypothesis 3 (Leader Mimicry). A lead post's offensive speech increases the likelihood of a replying comment containing offensive speech in the same discussion thread.

Hypothesis 4 (Peer Mimicry). A replying comment's offensive speech increases the likelihood of a later comment containing offensive speech.

3. Context of the study

This study focuses on Hong Kong's digital space between 2014 and 2018. Hong Kong used to take pride in its political plurality, moderation, pragmatism, and economic affluence, but the heightened tension with Mainland China in the recent decade has posed a range of political, economic, and governance challenges. The 2014 Umbrella Movement is the first large-scale social movement that manifested the conflict between the two regions, ensuing many other protests and movements in Hong Kong ([Stacey, 2018](#)).

The post-UM period has reflected a multilayered sociopolitical conflict between Mainland China and Hong Kong. The movement has left a deep mark on the minds of Hong Kong citizens, whose desire for localism and autonomy from Mainland China has hardened even further. Socio-economic and cultural tensions have also been substantial. For example, the Yuen Long Retrocession in 2015 showcased Hong Kong citizens' negative perception of Mainland Chinese parallel traders as the main cause for skyrocketing local product prices. Culturally, student protests at higher institutions revealed the resistance of local students to the cultural influence of Mainland China. In a nutshell, the post-UM period witnessed the rise of localism ([Kaeding, 2017](#)) and continued activism ([Cheng, 2016](#)), which intensified Hong Kongers' hostility toward Mainland Chinese.

The sociopolitical tension seems to have become a defining character

of post-UM digital culture (Song & Wu, 2018). For example, an antagonistic slang “Chee-na (also called Cina or Shina; 支那)” has circulated ubiquitously across major online forums to mock Mainland China and Chinese people as well as the tendency of “mainlandization”. The spillover influence of the economic rise of Mainland China has prompted Hong Kongers’ resistance to further integration. The term “Chee-na” was originally a neutral term yet evolved into a derogatory word after the Second Sino-Japanese War. With a deeply humiliating connotation of Japanese invasion and war crimes, it has been a taboo word until recently when some protesters reclaimed the term to demean Mainland China’s political system and social order. In the first post-UM 2016 Legislative Council election, two localist politicians were deprived of their election seats because of their use of “Chee-na” in their oaths. The reinvented semantic of “Chee-na” exemplifies the mainstreaming of divisive vernacular culture in the current Hong Kong digital space. In addition to examining the social contagion mechanisms of political swearing as a whole, this study also conducted an exemplary case study by testing the contagion of “Chee-na” as a specific language indicator for supplementary analyses.

4. Method

4.1. Data

Using the keywords “Mainland China” (大陆) or “Inland China” (内陆), we developed a Python programme to retrieve all the lead posts related and replying comments published in HKGolden’s current affairs section from September 28, 2014, to August 31, 2018. We retained posts that had at least one keyword appearing in the post title or the post content, and retrieved the corresponding conversational threads, including replying comments. Eventually, we obtained 28,506 posts and 1,126,455 replying comments made by a total of 35,116 users. We also retrieved identified metadata associated with posts, replying comments, and users.

4.2. Measurements

4.2.1. Dependent variables

Replying Comment’s Swearing (Use of “Chee-na”). First, we classified whether a replying comment included political swearing. We manually constructed a dictionary for swearing words or phrases that have been commonly contained in online political comments. We then constructed a binary variable (*reply_swearing*), where a value of “1” indicated that the focal replying comment included at least one of the words or phrases in the dictionary and “0” if otherwise. Similarly, our second dependent variable (*reply_Chee-na*) was whether a replying comment included the word “Chee-na,” where a value of “1” indicated that this reply contained the word “Chee-na” and “0” if otherwise.

4.2.2. Contagion mechanisms

Generalized Reciprocity. To quantify the generalized reciprocity effect, we counted the number of swearing comments that called out the author of the focal comment while excluding those quoted back directly by the author. We counted only the comments that came before the focal comment was made by considering two temporal windows: up to one month (*genrecip_month*) and one week (*genrecip_week*) before the focal comment was made. On average, 1.35 swearing comments called out (quoted) a target author for a month and 0.79 swearing comments for a week.

We computed *genrecip_month* and *genrecip_week* across the entire dataset, considering the possibility that the focal author was targeted in one discussion thread while reciprocating in a different thread. That said, the contagion effect may be more obvious within the same discussion thread context. Accordingly, we constructed another measure of generalized reciprocity within the immediate discuss thread context (*genrecip_thread*). Like *genrecip_month* and *genrecip_week*, this variable

was differentiated from the direct reciprocity effect by excluding comments that the focal author quoted back. The average of *genrecip_thread* was 2.21.

Applying the same approach, we constructed generalized reciprocity measures for “Chee-na.” The average scores for *genrecip_month*, *genrecip_week*, and *genrecip_thread* of a focal comment for “Chee-na” were 0.23, 0.14, and 0.40, respectively.

Direct Reciprocity. To quantify direct reciprocity effect (*direct_recip*), for the users that the author quoted in the focal comment, we counted the total number of preceding swearing (or “Chee-na”) comments from these users that quoted the author of the focal comment. This interaction could happen not only within the same thread but also when they met in another discussion thread. Accordingly, we examined all discussion threads and all time periods (before the focal comment was made) to measure direct reciprocity. An average score of *direct_recip* was 0.34 for the entire swearing comments, and 0.06 for the word “Chee-na.”

Leader Mimicry. A binary variable (*leader_swear*) was constructed to indicate whether an initial post that led a discussion thread contained political swearing (or the word “Chee-na”).

Peer Mimicry. Peer mimicry effect was measured as the percentage of comments within the same discussion thread that contained swearing (or the word of “Chee-na”) prior to the publication of a focal comment (*peer_swear*). We excluded preceding comments that quoted the focal user to avoid overwriting the reciprocity-related measures.

4.2.3. Control variables

The characteristics of a lead post can affect the dynamics of ensuing interactions. Accordingly, we controlled for several characteristics of a lead post. First, we quantified the emotional and cognitive elements of the post using TextMind. TextMind is a Chinese language text analysis system similar to the widely used software Linguistic and Word Count (LIWC), which calculates the percentage of words in a given text that belong to predefined linguistic categories (Tausczik & Pennebaker, 2010). We referred to the categories of positive emotion, negative emotion, anger, sadness, and cognition from the outputs that TextMind generated. Second, we controlled for a lead post’s topical nature (= politics) by applying a parallelized Latent Dirichlet Allocation algorithm-based topic modeling using a python package *gensim*. From iterative trials of the standard procedures of topic modeling, we concluded the 50 topics model to be the most interpretable model. Among the 50 topics, eight were political topics. We identified a political post if the most prominent topic of the post was one of these eight political topics.

4.3. Analysis plan

Linear probability models⁴ were used to analyze how contagion mechanisms are associated with the likelihood of an ensuing swearing comment. We separately tested each generalized reciprocity variable (i. e., *genrecip_month*, *genrecip_week*, and *genrecip_thread*) to avoid the multicollinearity problem, while comparing the results for robustness check (Table 1).

The unit of analysis was a comment. Each comment belonged to a user and a lead post. Given the panel nature of the data structure, we estimated the models by adding user-fixed effects, post-level characteristics (i.e., control variables, the number of replying comments a post engendered and reputation-indicative votes a post received), and temporal effects (i.e., time trend and seasonality, which indicated in which

⁴ The logistic regression model is another widely used model for a binary dependent variable. We chose linear probability modeling because logistic regression models failed to converge with our data. While not converged, logistic regression modeling results were consistent with the linear probability model.

Table 1
Correlations and summary statistics for key variables.

Variable	<i>M</i>	<i>SD</i>	<i>min</i>	<i>max</i>	1	2	3	4	5	6
Overall Swearing										
1. Reply_swear	0.18	0.39	0.00	1.00						
2. Leader_swear	0.48	0.50	0.00	1.00	0.03***					
3. Peer_swear (log)	−1.58	0.47	−2.35	−0.30	0.14***	0.10***				
4. Genrecip_month (log)	−2.13	0.74	−2.35	1.70	0.05***	0.04***	0.07***			
5. Genrecip_week (log)	−2.20	0.61	−2.35	1.61	0.05***	0.03***	0.07***	0.83***		
6. Genrecip_thread (log)	−2.13	0.75	−2.35	1.82	0.05***	0.06***	0.09***	0.75***	0.69***	
7. Direct_recip (log)	−2.26	0.47	−2.35	1.61	0.04***	0.02***	0.04***	0.18***	0.16***	0.22***
The Use of “Chee-na”										
1. Reply_Chee-na	0.03	0.18	0.00	1.00						
2. Leader_swear	0.02	0.14	0.00	1.00	0.04***					
3. Peer_swear (log)	−2.34	0.06	−2.35	−0.67	0.05***	0.07***				
4. Genrecip_month (log)	−2.30	0.38	−2.35	1.53	0.02***	0.02***	0.52***			
5. Genrecip_week (log)	−2.32	0.31	−2.35	1.53	0.02***	0.02***	0.49***	0.80***		
6. Genrecip_thread (log)	−2.30	0.40	−2.35	1.67	0.02***	0.03***	0.67***	0.76***	0.68***	
7. Direct_recip (log)	−2.33	0.22	−2.35	1.58	0.02***	0.02***	0.26***	0.07***	0.06***	0.09***

Notes. *M* and *SD* represent mean and standard deviation. *min* and *max* represent minimum and maximum values. All the correlations were tested at the comment level. Summary statistics for *Leader_swear* are at the post level. All the other summary statistics were at the reply level. *** $p < 0.001$.

month the focal comment was made). We took the same variables and modeling approaches to examine hypotheses with regard to the occurrence of “Chee-na.”

5. Results

5.1. Descriptive results

A preliminary review of data showed that posting activities in HKGolden remained active in the post-UM period, suggesting continuing interests and controversy surrounding Mainland China-related affairs. The number of comments relative to the number of posts surged during the fourth quarter of 2017 and later (Fig. 2), during which a series of incidents manifested growing interference from Mainland China directed at the economy and culture of Hong Kong, as described in the timeline of major events (Fig. 3).

Among 1,067,382 replying comments, 18.24% ($N = 194,641$) were identified as containing at least one of the political swearing words or phrases in the dictionary we constructed, and 3.5% ($N = 36,840$) were identified to contain the word “Chee-na.” Among 27,922 lead posts, 48.22% ($N = 13,463$) were identified as containing swearing, and 1.93% ($N = 540$) were identified as including the use of “Chee-na.” Contagion measures showed heavy skewness. We thus transformed the variables by taking the logarithm twice in the later analysis.⁵ Particularly, the use of “Chee-na” showed a sharp peak in October 2016 when politicians’ controversial oath-taking and disqualification began (Fig. 4).

Also compared is the proportion of swearing comments between the discussion threads that began with a swearing lead post and those with a non-swearing lead post. The trend showed a higher proportion of swearing comments with a lead swearing post across all time periods. The comparison for the use of “Chee-na” showed a similar yet starker pattern, again with a peak in December 2016 (Fig. 5).

5.2. Hypotheses tests: social contagion effects

Linear probability models assessed the effects of contagion mechanisms on the occurrence of swearing in a comment. Robust standard errors and standardized coefficients were reported in parentheses and square brackets, respectively (Tables 2 and 3). Because predictors were of different scales, standardized coefficients were useful in comparing the relative effects of different mechanisms.

Hypothesis 1 (H1) addressed the effect of generalized reciprocity on

offensive speech. For both overall swearing and the use of “Chee-na,” the results indicated that generalized reciprocity variables were significantly associated with a focal comment’s likelihood of swearing, supporting H1. That is, the more a focal user was targeted by swearing comments (by the reference of “Chee-na”), the more likely the focal user made a swearing comment (with “Chee-na”) to someone else as well.

Hypothesis 2 (H2) focused on the direct reciprocity effect. For both overall swearing and the use of “Chee-na,” the results indicated a significant positive association between the direct reciprocity variable and a focal comment’s likelihood of swearing, supporting H2. In other words, if the focal user was targeted by a swearing comment, the focal user was more likely to swear back at the instigator.

Hypothesis 3 (H3) inquired whether the swearing in the lead post contributed to engendering swearing in subsequent comments. For both overall swearing and the use of “Chee-na,” the results showed the positive significant association of a lead post’s swearing with a replying comment’s swearing, supporting the leader mimicry effect.

Lastly, **Hypothesis 4 (H4)** suggested the effect of peer mimicry, that is, the contagion of swearing from other preceding comments to the focal comment. Again, the results showed significant peer mimicry effects for both overall swearing uses and the use of “Chee-na.” The swearing of earlier comments was positively associated with the likelihood of a later comment’s swearing, supporting H4.

While the results showed all mechanisms were in effect, the comparison of standardized coefficients suggested that *peer mimicry* effects explained the swearing occurrence the most ($\beta \approx 0.09$), followed by direct reciprocity ($\beta \approx 0.03$), generalized reciprocity ($\beta \approx 0.02$), and leader mimicry ($\beta \approx 0.017$). In the case of “Chee-na,” peer mimicry ($\beta \approx 0.03$) and leader mimicry effects ($\beta \approx 0.03$) were larger than generalized and direct reciprocity ($\beta \approx 0.017$). The effect of direct reciprocity was significant yet minimal in the case of “Chee-na” ($\beta \approx 0.003$).

6. Discussion

Despite increased Mainland-Hong Kong economic and social integration in the recent decade, the dislocating pace of change has created rising insecurities for Hong Kong society who assign blame to Mainland China for many of their frustrations. Expressing resistance and hostility against the growing influence of Mainland China seems to have become a norm in Hong Kong’s digital spaces. With political swearing as an exemplar of offensive speech, this study took a social interactional approach to understand how individuals’ offensive comments became a collective speech style.

While the literature of digital incivility has primarily focused on individual user attributes (e.g., Rains et al., 2017) and content

⁵ We transformed the variables by applying $\log(\log(x+1.1))$.

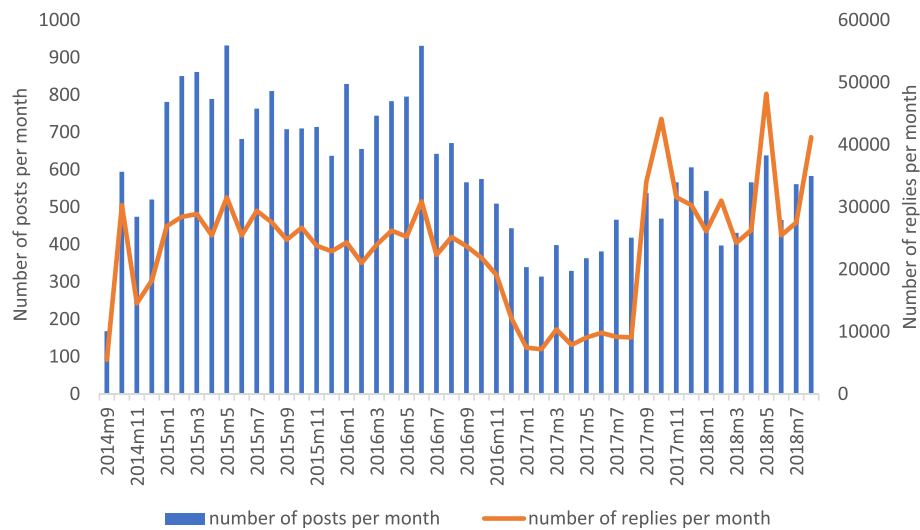


Fig. 2. Number of lead posts and replying comments per month.

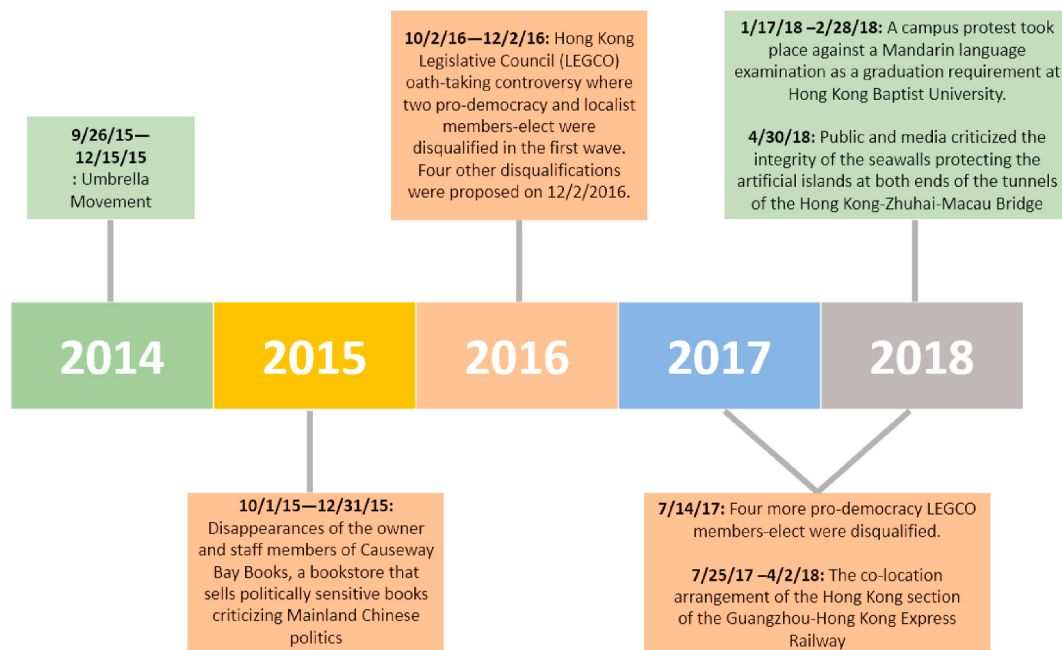


Fig. 3. Timeline of major political events between 2014 and 2018.

characteristics (e.g., Ng et al., 2020), the current study paid attention to social network dynamics. Social networks compose not only dyadic relationships but also clique-level small groups and the crowd-level of a large group. An online forum is likewise a multilayered social system. While this study could not capture all possible network dynamics, it examined a few noteworthy interaction mechanisms by proposing four mechanisms of social contagion: generalized reciprocity, direct reciprocity, leader mimicry, and peer mimicry. Whereas direct reciprocity occurs in a dyad, leader mimicry and peer mimicry do not necessarily entail direct interactions. Furthermore, generalized reciprocity occurs in an open network context.

Overall, our findings suggest that offensive speech is indeed contagious. Political swearing influenced not only the targeted users but also bystanders' commenting behaviors. Most interestingly, among all mechanisms, we found that peer mimicry showed the largest effect on the occurrence of political swearing. This was the case for both overall swearing and the use of the individual swearword "Chee-na." Given that

most people may read others' offensive speech rather than being directly targeted by it, the significant mimicry effects suggest a spiraling potential of individuals' swearing utterances into a collective discursive culture. Social learning may occur in this process: A third person may observe hostile interactions between instigators and targets, and in turn take it as an acceptable speech act and model it in his or her future interactions. Community members learn normative reactions by observing their peers, and one is even more likely to synchronize if multiple peers, rather than a single peer, exhibit the same behavior.

Social contagion can become more complex in an online context than offline because online social transmission entails multiple communication carriers simultaneously. Online discussion forums provide an environment in which exposure to normative behaviors, interactional synchrony, and social reinforcement take place. Community members learn and adapt their speech styles by reciprocal interaction and observation. In addition to the mimicry effects, direct and generalized reciprocity were also significant contributors: A targeted user may

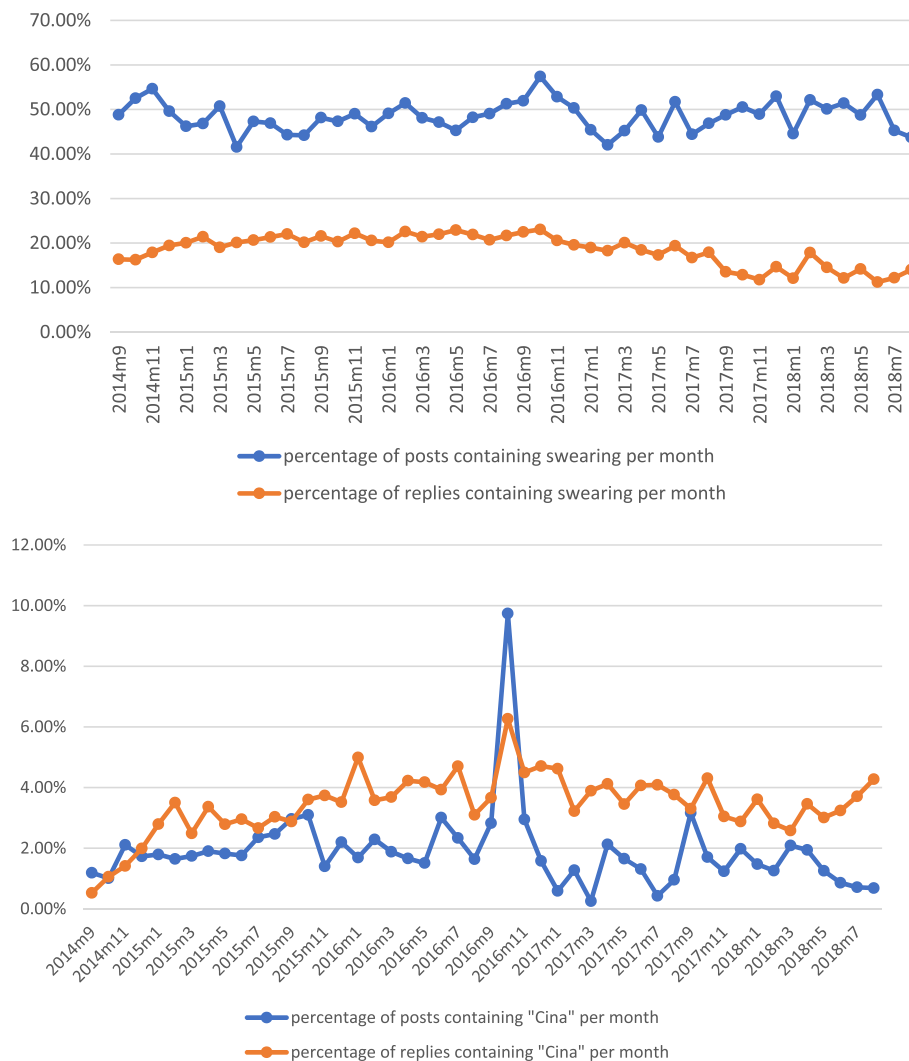


Fig. 4. Percentage of posts and replies with political swearing. The upper graph considers all swearwords and phrases; the bottom graph considers the occurrence of “Chee-na”.

displace his or her irritated mood by directing offensive speech not only back at the instigator but also at unrelated users. We found evidence of both mimicry and reciprocity effects, confirming that all hypothesized mechanisms contributed to the incivility spiral.

While leader mimicry (i.e., imitating the speech style of the initial post) explained the smallest variance of overall swearing speech, it played a more important role in the context of adopting the word “Chee-na.” The difference in leader mimicry between the overall swearing and the word “Chee-na” reiterates the importance of understanding the cultural meaning of an utterance. The larger effect of leader mimicry on the adoption of “Chee-na” alludes to the situation in which the word could have served as a signifier of group identity. Rather than being used as an emotional utterance to attack other users within the community, this word appears to have evolved into a linguistic meme that denigrates *outsiders* of the online community. Posts and comments in our dataset notably showed the use of “Chee-na” in reference to Mainland Chinese tourists. From the social learning perspective, we interpret that the adoption of the word “Chee-na” has become a descriptive norm, particularly being initiated by the lead posters and then accepted by replying commenters (Zhou & Fang, 2015). The findings with regard to “Chee-na” resonate with Graeff’s (2015) discussion on meme speech acts that demarcate the in-group status of a networked public: The online public – initially a network of strangers – demonstrates and negotiates in-group belonging by adopting memes and other shared speech acts.

Mimicry thus serves as an evolved tool for “social glue”, which allows people to quickly and unconsciously communicate rapport with other in-group members in the discussion community. The implicit formation of an “us” versus “them” that undergirds interactional synchrony is thus situated in the “memetic logic” that manifests in- and out-group distinctions (Tuters & Hagen, 2020, p. 2225).

7. Conclusion

The purpose of this study aims to investigate how Hong Kong’s networked culture has become increasingly tolerant of (and even normalizing) offensive political vernaculars. Drawn from anti-social contagion literature, this study investigates the social interactional processes underlying the incivility spiral of mediated political conversations. Specifically, this study examines four social interactional mechanisms underlying a user’s adoption of political swearing: generalized reciprocity, direct reciprocity, leader-mimicry, and peer-mimicry. Findings show that peer-mimicry contributes to the contagion process the most, followed by generalized reciprocity and direct mimicry. Norms can be created and solidified through various social mechanisms embedded in discussion dynamics. This study’s findings confirm that social norms of an online community are emergent, arise through the accumulation of micro-social interactions, and respond to the shifting discussion dynamics within a community. Understanding the

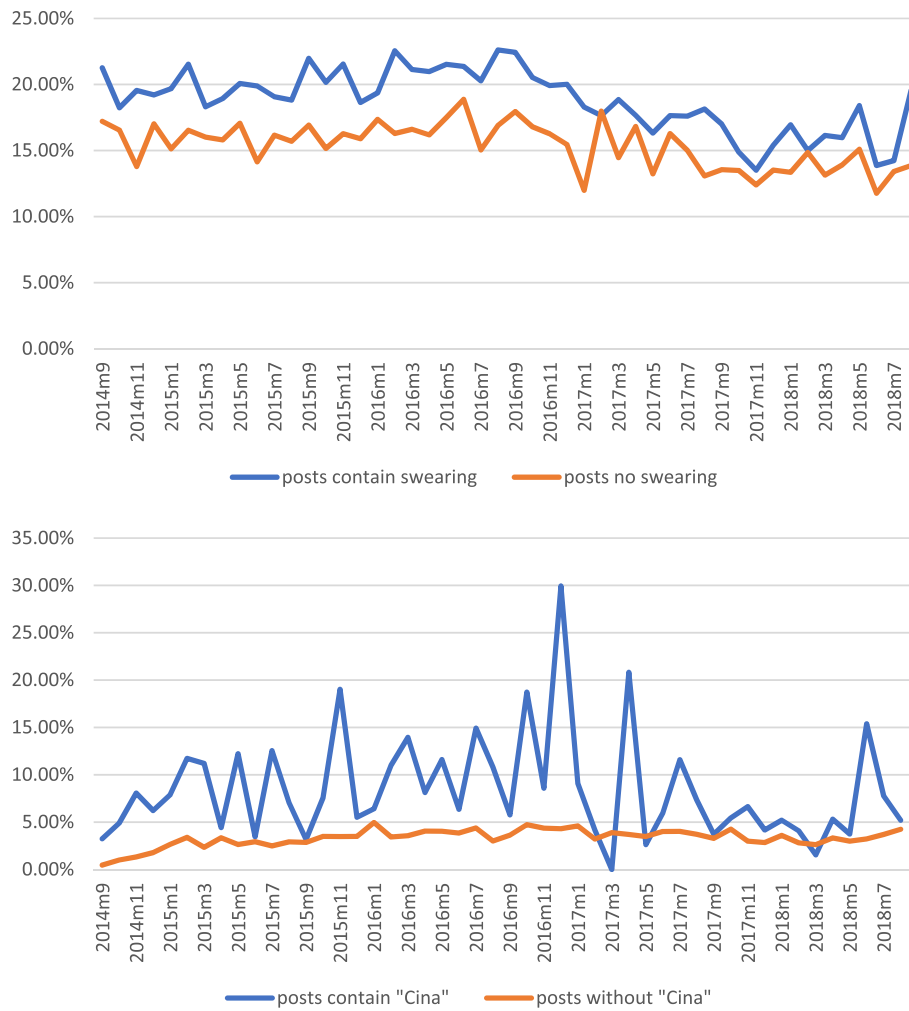


Fig. 5. Percentage of swearing replies over time by lead post type. The upper graph considers all swearwords and phrases; the bottom graph considers the occurrence of “Chee-na”.

Table 2

Linear probability models predicting contagion mechanism effects on occurrence of political swearing in replies ($N = 34,730$).

Variables	M1	M2	M3
Leader_swear	0.0129*** (0.00140) [0.0167]	0.0129*** (0.00140) [0.0167]	0.0128*** (0.00139) [0.0165]
Peer_swear (log)	0.0766*** (0.00203) [0.0924]	0.0767*** (0.00202) [0.0925]	0.0762*** (0.00200) [0.0919]
Genrecip_month (log)	0.0112*** (0.00193) [0.0214]		
Genrecip_week (log)		0.0127*** (0.00263) [0.0201]	
Genrecip_thread (log)			0.0134*** (0.00215) [0.0262]
Direct_recip (log)	0.0246*** (0.00369) [0.0305]	0.0249*** (0.00372) [0.0308]	0.0238*** (0.00371) [0.0295]
Observations	1,067,382	1,067,382	1,067,382
R-squared	0.014	0.014	0.014

Notes. Robust standard error in parentheses and standardized coefficients in square brackets.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Control variables were not reported due to space limitations. Full results are available upon request.

Table 3

Linear probability models predicting contagion mechanism effects on occurrence of “Chee-na” in replies ($N = 34,730$).

Variables	M4	M5	M6	M7
Leader_swear	0.0391*** (0.00295) [0.0317]	0.0411*** (0.00299) [0.0333]	0.0410*** (0.00299) [0.0332]	0.0409*** (0.00299) [0.0332]
Peer_swear (log)	0.0998*** (0.0164) [0.0338]			
Genrecip_month (log)		0.00768*** (0.00233) [0.0161]		
Genrecip_week (log)			0.0104** (0.00321) [0.0175]	
Genrecip_thread (log)				0.00845** (0.00265) [0.0183]
Direct_recip (log)	0.00311 (0.00269) [0.00368]	0.00922*** (0.00256) [0.0109]	0.00923*** (0.00257) [0.0109]	0.00902*** (0.00256) [0.0107]
Observations	1,067,382	1,067,382	1,067,382	1,067,382
R-squared	0.003	0.003	0.003	0.003

Notes. Robust standard error in parentheses and standardized coefficients in square brackets.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. Control variables were not reported due to the space limitation. Full results are available upon request.

mechanisms of user-to-user interactions is a prerequisite toward understanding online conversational culture.

To our knowledge, this study is the first effort that quantifies social contagion effects underlying the emergence of offensive culture online. That said, the study has a few limitations. First, some offensive speech may not include a swear word or phrase. Also, the swearing dictionary we created is not an exhaustive collection of all relevant vernaculars. Digital users are creative collectives that constantly remix, reinvent, and play with languages. Political swearing is not an exception. Second, it is beyond the scope of the current study to address a causal relationship between offensive digital culture and the political situation of the city. One could ask whether the spread of offensive speech in Hong Kong's digital spaces is unique to the city's geopolitical context or a snapshot of the universal digital culture and a generalizable trend beyond Hong Kong's specific context. Future research can address such questions by gaining more qualitative insights through interviews, field observations, and comparative analyses across different sociopolitical contexts.

Credit author statement

Yunya Song: Conceptualization, Methodology, Formal Analysis, Investigation, Data curation, Writing – original draft preparation. **Qinyun Lin:** Methodology, Formal Analysis, Data curation, Writing – original draft preparation. **K. Hazel Kwon:** Conceptualization, Methodology, Formal Analysis, Investigation, Writing – Reviewing and Editing. **Christine H.Y. Choy:** Investigation, Writing – original draft preparation. **Ran Xu:** Validation, Writing – Reviewing and Editing.

Acknowledgements

The research reported in this article was supported by a General Research Fund grant (HKBK 12605520) of the Research Grant Council of Hong Kong, the Interdisciplinary Research Clusters Matching Scheme (IRCMS/19-20/D04) and the Initiation Grant for Faculty Niche Research Areas (RC-FNRA-IG-18-19-04) of Hong Kong Baptist University. The third author's contribution was supported by the National Science Foundation (Award ID 2027387).

References

- Alonso, M., & Aiken, M. (2004). Flaming in electronic communication. *Decision Support Systems*, 36(3), 205–213. [https://doi.org/10.1016/S0167-9236\(02\)00190-2](https://doi.org/10.1016/S0167-9236(02)00190-2)
- Arthur, M. M. L. (2013). Emergent norm theory. In D. A. Snow (Ed.), *The Wiley-Blackwell encyclopedia of social and political movements*. Blackwell Publishing Ltd. <https://doi.org/10.1002/9781405165518.wboese033.pub2>
- Ballard, M. E., & Welch, K. M. (2017). Virtual warfare: Cyberbullying and cyber-victimization in MMOG play. *Games and Culture*, 12(5), 466–491. <https://doi.org/10.1177/1555412015592473>
- Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, 28(3), 12–29. <https://doi.org/10.1111/j.1460-2466.1978.tb01621.x>
- Bass, F. M. (2004). Comments on “a new product growth for model consumer durables the bass model”. *Management Science*, 50(12, supplement), 1833–1840. <https://doi.org/10.1287/mnsc.1040.0300>
- Bastampillai, T., Allison, S., & Chan, S. (2013). Is depression contagious? The importance of social networks and the implications of contagion theory. *Australian and New Zealand Journal of Psychiatry*, 47(4), 299–303. <https://doi.org/10.1177/0004867412471437>
- Bies, R. J., & Moag, J. S. (1986). Interactional justice: Communications criteria of fairness. In R. Lewicki, M. Bazerman, & B. Sheppard (Eds.), *Research on negotiation in organization* (Vol. 1, pp. 43–55). JAI Press.
- Brown, B. B., & Klute, C. (2003). Friendships, cliques, and crowds. In G. R. Adams, & M. D. Berzonsky (Eds.), *Blackwell handbook of adolescence* (pp. 330–348). Blackwell Publishing.
- Chandrasekharan, E., Samory, M., Jhaver, S., Charvat, H., Bruckman, A., Lampe, C., ... Gilbert, E. (2018). The internet's hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro Scales. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–25. <https://doi.org/10.1145/3274301>
- Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 41, pp. 219–274). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)00405-X](https://doi.org/10.1016/S0065-2601(08)00405-X)
- Cheng, E. (2016). Street politics in a hybrid regime: The diffusion of political activism in post-colonial Hong Kong. *The China Quarterly*, 226, 383–406. <https://doi.org/10.1017/S0305741016000394>
- Cheng, J., Danescu-Niculescu-Mizil, C., Leskovec, J., & Bernstein, M. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. *American Scientist*, 105(3), 152–155. <https://doi.org/10.1511/2017.126.152>
- Cheung, R. (2019, August 8). *Hong Kong protest art: Meet the student leading the defiant design team*. South China Morning Post. <https://www.scmp.com/magazines/post-magazine/long-reads/article/3024096/hong-kong-protest-art-meet-student-leading>
- Cho, D., & Kwon, K. H. (2015). The impacts of identity verification and disclosure of social cues on flaming in online user comments. *Computers in Human Behavior*, 51, 363–372. <https://doi.org/10.1016/j.chb.2015.04.046>
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A focus theory of normative conduct: A theoretical refinement and reevaluation of the role of norms in human behavior. *Advances in Experimental Social Psychology*, 24, 201–234. [https://doi.org/10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10.1111/jcom.12104>
- Coyne, S. M., Stockdale, L. A., Nelson, D. A., & Fraser, A. (2011). Profanity in media associated with attitudes and behavior regarding profanity use and aggression. *Pediatrics*, 128(5), 867–872. <https://doi.org/10.1542/peds.2011-1062>
- De Dreu, C. K. W., & van Knippenberg, D. (2005). The possessive self as a barrier to conflict resolution: Effects of mere ownership, process accountability, and self-concept clarity on competitive cognitions and behavior. *Journal of Personality and Social Psychology*, 89(3), 345–357. <https://doi.org/10.1037/0022-3514.89.3.345>
- Fearn, N. (2017, January 30). *Twitter and the scourge of cyberbullying*. <https://www.idgconnet.com/idgconnect/analysis-review/1005120/twitter-scourge-cyberbullying>
- Frenda, S. (2018). The role of sarcasm in hate speech. A multilingual perspective. In *e Doctoral symposium of the XXXIV International conference of the Spanish society for natural language processing (SEPLN 2018)* (pp. 13–17).
- Graeff, E. (2015). *Binders full of election memes: Participatory culture invades the 2012 U.S. election*. Civic Media Project. <http://civicmediaproject.org/works/civic-media-project/binders-full-of-election-memes-participatory-culture-invades-the-2012-us-election>
- Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6), 1420–1443. <https://doi.org/10.1086/226707>
- Herring, S., Job-Sluder, K., Scheckler, R., & Barab, S. (2002). Searching for safety online: Managing “trolling” in a feminist forum. *The Information Society*, 18(5), 371–384. <https://doi.org/10.1080/01972240290108186>
- 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>
- Hodas, N. O., & Lerman, K. (2014). The simple rules of social contagion. *Scientific Reports*, 4(1), 1–7. <https://doi.org/10.1038/srep04343>
- Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). How could you think that?!: Understanding intentions to engage in political flaming. *New Media & Society*, 17(8), 1201–1219. <https://doi.org/10.1177/1461444814522947>
- Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science*, 4(2), 153–161. <https://doi.org/10.1111/j.1745-6924.2009.01115.x>
- Kaeding, M. (2017). The rise of “Localism” in Hong Kong. *Journal of Democracy*, 28(1), 157–171. <https://doi.org/10.1353/jod.2017.0013>
- Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137–146). <https://doi.org/10.1145/956750.956769>
- Kim, J. E., & Tsvetkova, M. (2020). *Large-scale network analysis reveals cheating spreads through victimization and observation*. arXiv:2003.11139.
- Kwon, K. H., & Cho, D. (2017). Swearing effects on citizen-to-citizen commenting online: A large-scale exploration of political versus nonpolitical online news sites. *Social Science Computer Review*, 35(1), 84–102. <https://doi.org/10.1177/0894439315602664>
- Kwon, K. H., & Gruzd, A. (2017). Is offensive commenting contagious online? Examining public vs interpersonal swearing in response to donald trump's YouTube campaign videos. *Internet Research*, 27(4), 991–1010. <https://doi.org/10.1108/IntR-02-2017-007>
- Kwon, K. H., Stefanone, M. A., & Barnett, G. A. (2014). Social network influence on online behavioral choices: Exploring group formation on social network sites. *American Behavioral Scientist*, 58(10), 1345–1360. <https://doi.org/10.1177/0002764214527092>
- Lampe, C., Zube, P., Lee, J., Park, C. H., & Johnston, E. (2014). Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, 31(2), 317–326. <https://doi.org/10.1016/j.giq.2013.11.005>
- Larson, J. J., Whitton, S. W., Hauser, S. T., & Allen, J. P. (2007). Being close and being social: Peer ratings of distinct aspects of young adult social competence. *Journal of Personality Assessment*, 89(2), 136–148. <https://doi.org/10.1080/00223890701468501>
- Lee, F. L. (2018). Internet alternative media, movement experience, and radicalism: The case of post-Umbrella Movement Hong Kong. *Social Movement Studies*, 17(2), 219–233. <https://doi.org/10.1080/14742837.2017.1404448>
- Lee, S. H., & Kim, H. W. (2015). Why people post benevolent and malicious comments online. *Communications of the ACM*, 58(11), 74–79. <https://doi.org/10.1145/2739042>
- March, E. (2019). Psychopathy, sadism, empathy, and the motivation to cause harm: New evidence confirms malevolent nature of the Internet Troll. *Personality and Individual Differences*, 141, 133–137. <https://doi.org/10.1016/j.paid.2019.01.001>

- Marsden, P. (1998). Memetics and social contagion: Two sides of the same coin? *Journal of Memetics Evolutionary Models of Information Transmission*, 2(2), 171–185.
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on youtube. *Computers in Human Behavior*, 26(6), 1536–1546. <https://doi.org/10.1016/j.chb.2010.05.023>
- Ng, Y. L., Song, Y., Kwon, K. H., & Huang, Y. (2020). Toward an integrative model for online incivility research: A review and synthesis of empirical studies on the antecedents and consequences of uncivil discussions online. *Telematics and Informatics*, 47, Article 101323. <https://doi.org/10.1016/j.tele.2019.101323>
- Ortiz, S. M. (2019). “You can say I got desensitized to it”: How men of color cope with everyday racism in online gaming. *Sociological Perspectives*, 62(4), 572–588. <https://doi.org/10.1177/0731121419837588>
- Patterson, G. R. (1982). *Coercive family processes*. Castalia.
- 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>
- Reicher, S. D. (1996). “The Battle of Westminster”: Developing the social identity model of crowd behaviour in order to explain the initiation and development of collective conflict. *European Journal of Social Psychology*, 26(1), 115–134. [https://doi.org/10.1002/\(SICI\)1099-0992\(199601\)26:1<115::AID-EJSP740>3.0.CO;2-Z](https://doi.org/10.1002/(SICI)1099-0992(199601)26:1<115::AID-EJSP740>3.0.CO;2-Z)
- Reicher, S. D., & Stott, C. (2011). *Mad mobs and englishmen? Myths and realities of the 2011 riots*. Constable & Robinson.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). Free Press.
- Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web* (pp. 695–704). <https://doi.org/10.1145/1963405.1963503>
- Rossini, P., Maia, R., & Maia, R. C. (2021). Characterizing disagreement in online political talk: Examining incivility and opinion expression on news websites and Facebook in Brazil. *Journal of Deliberative Democracy*, 17(1), 90–104. <https://doi.org/10.16997/10.16997/jdd.967>
- Rowe, I. (2014). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, Communication & Society*, 18(2), 121–138. <https://doi.org/10.1080/1369118X.2014.940365>
- Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4), 296–320. https://doi.org/10.1207/S15327957PSPR0504_2
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755–1758. <https://doi.org/10.1126/science.1082976>
- Seering, J., Kraut, R. E., & Dabbish, L. (2017). Shaping pro and anti-social behavior on twitch through moderation and example-setting. In *Proceedings of the ACM conference on computer supported Cooperative Work* (pp. 111–125). CSCW. <https://doi.org/10.1145/2998181.2998277>
- Shachaf, P., & Hara, N. (2010). Beyond vandalism: Wikipedia trolls. *Journal of Information Science*, 36(3), 357–370. <https://doi.org/10.1177/0165551510365390>
- Shmargad, Y., Coe, K., Kenski, K., & Rains, S. A. (2021). Social norms and the dynamics of online incivility. *Social Science Computer Review*, 1–19. <https://doi.org/10.1177/0894439320985527>
- Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The roles of distributive, procedural, and interactional justice. *Journal of Applied Psychology*, 82(3), 434–443. <https://doi.org/10.1037/0021-9010.82.3.434>
- Song, Y., Kwon, K. H., Xu, J., Huang, X., & Li, S. (2021). Curbing profanity online: A network-based diffusion analysis of profane speech on Chinese social media. *New Media & Society*, 23(5), 982–1003. <https://doi.org/10.1177/1461444820905068>
- Song, Y., & Wu, Y. (2018). Tracking the viral spread of incivility on social networking sites: The case of cursing in online discussions of Hong Kong–Mainland China conflict. *Communication and the Public*, 3(1), 46–61.
- Stacey, E. (2018). Networked protests: A review of social movement literature and the Hong Kong Umbrella movement (2017). In S. Chhabra (Ed.), *Handbook of research on civic engagement and social change in contemporary society* (pp. 347–363). IGI Global.
- Su, A. (2019, June 2). Crackdown, arrests loom over Hong Kong as martyrdom becomes part of protest narratives. Los Angeles Times. <https://www.latimes.com/world/asia/la-fg-hong-kong-protests-violence-martyrdom-20190702-story.html>
- Suler, J. (2004). The online disinhibition effect. *CyberPsychol and Behavior*, 7(3), 321–326. <https://doi.org/10.1089/1094931041291295>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of qords: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Tedeschi, J. T., & Felson, R. B. (1994). *Violence, aggression, and coercive actions*. American Psychological Association. <https://doi.org/10.1037/10160-000>
- Toma, C. L. (2014). Counting on friends: Cues to perceived trustworthiness in Facebook profiles. In *Proceedings of the international conference on Weblogs and social media (ICWSM)*. Association for the Advancement of Artificial Intelligence Press.
- Tsvetkova, M., & Macy, M. W. (2015). The social contagion of antisocial behavior. *Sociological Science*, 2, 36–49. <https://doi.org/10.15195/v2.a4>
- Turner, R. H., & Killian, L. M. (1987). *Collective behavior* (3rd ed.). Prentice-Hall, Inc.
- Tuters, M., & Hagen, S. (2020). ((They))) rule: Memetic antagonism and nebulous othering on 4chan. *New Media & Society*, 22(12), 2218–2237. <https://doi.org/10.1177/1461444819888746>
- Valente, T. W. (1996). Social network thresholds in the diffusion of innovations. *Social Networks*, 18(1), 69–89. [https://doi.org/10.1016/0378-8733\(95\)00256-1](https://doi.org/10.1016/0378-8733(95)00256-1)
- Varjas, K., Talley, J., Meyers, J., Parris, L., & Cutts, H. (2010). High school students’ perceptions of motivations for cyberbullying: An exploratory study. *Western Journal of Emergency Medicine*, 11(3), 269–273.
- Ver Steeg, G., Ghosh, R., & Lerman, K. (2011). What stops social epidemics? In *Fifth international AAAI conference on Weblogs and social media*.
- Wilson, J. Q., & Kelling, G. L. (1982). Broken windows. *Atlantic Monthly*, 249(3), 29–38.
- Wong, R. Y., Cheung, C. M., & Xiao, B. (2018). Does gender matter in cyberbullying perpetration? An empirical investigation. *Computers in Human Behavior*, 79, 247–257. <https://doi.org/10.1016/j.chb.2017.10.022>
- Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, 10(1), 115–121. <https://doi.org/10.1089/cpb.2006.9984>
- Yokotani, K., & Takano, M. (2021). Social contagion of cyberbullying via online perpetrator and victim networks. *Computers in Human Behavior*, 119, Article 106719. <https://doi.org/10.1016/j.chb.2021.106719>
- Zhou, N., & Fang, X. Y. (2015). Beyond peer contagion: Unique and interactive effects of multiple peer influences on Internet addiction among Chinese adolescents. *Computers in Human Behavior*, 50, 231–238. <https://doi.org/10.1016/j.chb.2015.03.083>