Article

SM+S
social media + society

Social Media + Society
January-March 2023: 1–12
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20563051231157299
journals.sagepub.com/home/sms
SAGE

The Effects of Observer Expectations on Judgments of Anti-Asian Hate Tweets and Online Activism Response

Stephanie Tom Tong[1] and David C. DeAndrea[2]

Abstract

The rise of racial hate speech on social media has raised critical questions for scholars to explore. It is necessary to understand how outside observers passively evaluate (a) online racial hate speech posts on social media and (b) whether those evaluations are related to observers' subsequent behavior. This study explored how observers evaluate acts of majority-on-minority and minority-on-minority anti-Asian hate tweets on Twitter. In an experiment ($n = 196$) informed by expectancy violations theory, we tested how White observers evaluated anti-Asian tweets ostensibly posted by either a White or Black source. Analysis revealed a moderated-mediation pathway in which observers' political partisanship (Democrat/Republican) affected how they judged the ethnic prototypicality of White and Black sources of racial hate speech; these source prototypicality judgments were in turn associated with observers' judgments of tweet offensiveness and self-reported intentions to engage in online activism (i.e., signing an online petition). These results contribute to our understanding of outside observers' differential expectancies regarding online hate speech, and how those expectancies can affect perceptions of and reactions to acts of racism.

Keywords

Asian Americans, expectancy violations theory, interminority relations, online hate, performative allyship

Although no one agreed-upon scholarly definition exists, at its core, hate speech consists of hostile messages that "target a group, or an individual as they relate to a group" (Sellars, 2016, p. 25). Groups are often identified by salient characteristics—elements like religion, sexuality, gender, or race. This study focuses on *racial hate speech*, or hostile language that deliberately discriminates against individuals specifically because of their race or ethnicity (Bliuc et al., 2018). The features and affordances of popular social media platforms have changed the way contemporary racial hate speech is communicated and received online. What may have once been a face-to-face insult or microaggression aimed at one person or a small group of listeners has now morphed into a broader, more public online performance. In a setting like Twitter, for example, one tweet can reach an audience of thousands of outside (or "third-party") passive observers who are not members of the targeted group, but who nevertheless see such posts and make judgments about the offensiveness or appropriateness of a message and its source. Given that 53.34% of Americans report being exposed to online hate (Hawdon et al., 2017), understanding which factors influence outside observers' perceptions and responses to such content is extremely important.

In this way, rather than thinking of a hate message as a basic locutionary act in which words convey meanings, in online settings, racial hate messages function as powerful speech acts that transmit those meanings to a larger audience of passive observers in a way that can motivate subsequent responses (Searle, 1965). Indeed, research has documented associations between repeated viewing of hateful content online and a variety of harmful effects for outside observers. Although some may merely withdraw from public political conversation or debate after viewing online hate speech (Barnidge et al., 2019) for other observers, repeated exposure can result in desensitization to verbal acts of racial harassment and violence, increased feelings of prejudice and greater distancing from targeted outgroups (Soral et al., 2018). Sometimes, seeing online hate can inspire observers

[1]Wayne State University, USA
[2]The Ohio State University, USA

Corresponding Author:
Stephanie Tom Tong, Department of Communication, Wayne State University, 906 W. Warren Ave., 585 Manoogian Hall, Detroit, MI 48201, USA.
Email: stephanie.tong@wayne.edu

to produce similar content; Walther (in press) notes that when individuals make the move from passive observers to active perpetrators, they spread online hate in an effort to garner social approval and support from like-minded haters. On the other hand, sometimes viewing online hate can incite observers to defensive actions, such as counter-speech or online activism (Mathew et al., 2018; Meyers et al., 2020). Because observers' responses to racial hate speech can be so varied, the goal of the current study is to better understand what factors affect their reactions to offensive speech acts they see performed online.

In this study, we focus specifically on how White observers perceive and react to online racial hate because as members of the dominant majority, they are positioned to function as allies in the collective fight against racism (DeTurk, 2011). As such, their role as passive observers is a powerful one—in essence, this study asks the question if White observers "see something" online, will they "say something" or "do something" and if so, what? We center our investigation in the context of COVID-19 racial hate speech on Twitter, where anti-Asian tweets have risen an estimated 900% since the beginning of the pandemic (Gilbert, 2020). As frustration and fear has increased during the pandemic, many people have scapegoated Chinese Americans—and all Asian Americans, more broadly—as being the cause of the coronavirus (Li & Nicholson, 2021; Tong et al., 2022). In 2021, the Anti-Defamation League reported in the results of their annual survey of online hate and harassment that Asian American respondents "experienced the largest single year-over-year rise in severe online harassment in comparison to any other groups" (p. 6). Therefore, the pandemic provides an organic arena in which to study outside observers' reactions to anti-Asian online hate speech.

Informed by the expectancy violations framework, we experimentally test how variations in White observers' political partisanship interact with variations in the source's (i.e., offender's) race to influence observers' expectations regarding the source's ethnic prototypicality. In turn, we examine how observers' judgments of the source's ethnic prototypicality (a) are used to make subsequent evaluations about the offensiveness of the hate tweet, and (b) affect observers' intentions to engage in online activism—specifically, signing an online petition.

Observers' Expectations, Perceptions, and Evaluations of Offensive Speech Acts

As noted above, a single definition of hate speech is difficult to find. This is because by nature, speech acts are extremely flexible, and acts of online hate and incivility can vary considerably. Kenski and colleagues (2020) found significant variation in observers' evaluations of five kinds of online incivility, with name-calling and vulgar language rated as the most offensive. Because judgments regarding the potential offense of others' behavior is in the eye of the observer, one framework that helps explain how people form those judgments is expectancy violations theory (EVT; Burgoon & Hale, 1988; Burgoon & Walther, 1990). EVT posits that observers develop expectations about the appropriateness or normalcy of others' behavior. When a target's behavior violates their expectations, observers' arousal is stimulated, prompting them to make judgments about the target and the behavior. *Valence* of the violation and judgment of the target often parallel each other: If the observer perceives the behavioral violation as a negative thing, they will judge the target more negatively than if they had behaved according to expectations. On the other hand, if the violation was perceived as incongruous with expectations in a positive way, positive assessments are predicted to follow.

Expectancies are often formed at a sociological or group level—for example, an individual might hold gendered role expectations about the ways men and women "should" behave. Interestingly, the exact same behavior can be interpreted in different ways by different observers depending on the expectancies they hold. For example, imagine there are two people who are judging a scenario in which a woman manager is leading her team in the workplace. One observer finds that the manager's behavior violates their gender role expectations negatively and so judges the manager as "overly demanding." On the other hand, a different observer might view the manager's leadership style as a positive violation—as a display of confidence or power—leading to evaluations like "strong" and "ambitious." The expectancies that each observer holds can cast a long shadow over their final judgments of the behavior and its source.

Effect of Group-Level Expectancies on Judgments of Online Hate Speech

Although EVT was initially conceived of in the realm of nonverbal face-to-face interaction, it has also been applied to explain how people make sense of interactions in other environments, including in online spaces (e.g., Tong & Walther, 2015). In judging an anti-Asian tweet, an observer would likely evaluate the source of the tweet, as well as the tweet itself, when evaluating the extremity of potential effects. Following Walther (in press), we focus on racist tweets that are group-directed, or those in which a perpetrator expresses contempt for a particular racial or ethnic group, but "without implying any specific individuals and without targeting a specific person" (p. 6). This makes observers' *group-level expectations* especially relevant to the current study.

According to EVT, observers should judge a blatant act of racist hate speech as being more or less normative, anticipated, or offensive depending on the group-level expectations that are being activated during evaluation (Bettencourt et al., 2016). That is, observers often compare whether the target's behavior deviates from how they expect members of a salient group to behave. As in the example above, observers' expectations about gender roles were activated when judging the manager's leadership style. A related question is which group-level factors might affect observers' perceptions and evaluations in the context of online racial hate speech.

Factors Affecting Observers' Judgments of Hate Speech by Majority Sources

A review of past racial hate speech research reveals that most studies explore how *observers' ethnic group memberships* affect their judgments of racial hate speech messages that are depicted as being communicated by a White source (e.g., Cowan & Hodge, 1996; Cowan & Mettrick, 2002; Leets, 2001, 2003; Leets & Giles, 1997; Meyers et al., 2020). From these studies we know that racial minority and majority (White) observers' perceptions of racist messages communicated by White individuals can differ: In past experiments, Leets and Giles (1997) used hypothetical scenarios in which a White source was depicted communicating a directly racist (explicit language) or indirectly racist (ambiguous language) message to an Asian target. An interesting pattern emerged in which White observers found the directly racist messages to be more offensive than Asian American observers, while Asian American observers found the indirectly racist messages to be more offensive than White observers. Leets (1999, as cited in Leets, 2001) found similar patterns in which White observers, who read vignettes featuring a White source communicating explicitly racist messages to minority targets (Asian, Hispanic, and Black), evaluated the vignettes as more problematic compared to minority observers. Prior work also shows us how other contextual factors—observers' personal experience with racial harassment, speech act explicitness, and so on—can affect observers' judgments of racial hate speech.

Although insightful, this past research has almost exclusively used vignettes and scenarios that depict "classic" situations of "majority-on-minority" hate speech, in which a White perpetrator attacks a minority target. This setup reflects the "*prototypical expectancy*" in which "certain forms of discriminatory behavior serve as classic or 'best' examples of prejudice and discrimination" (Marti et al., 2000, p. 404; see also Baron et al., 1991). Consequently, it remains unknown what factors affect observers' judgments of other, non-White sources who communicate racial hate speech.

Salient Expectations Activated by the Message Source's Race

As a prototypically expected form of behavior, racist hate speech is often thought to be perpetrated by a majority (White) source against a minority target. In the current context, then, observers may judge anti-Asian COVID-related hate tweets from a White source as *ethnically prototypical*, which is defined as a behavior that is representative of, more expected, or more likely to be performed by members of the larger (majority White) group. Such an expectation would not be completely off-base: We know that most acts of online racist hate speech tend to be performed by members of White extremist groups (Bliuc et al., 2018; Costello et al., 2019; Daniels, 2017), but they are not the *only* perpetrators. Racist hate speech communicated by members of minority groups to other minorities can also occur—and although some research has examined acts of "cross-minority" or "minority-on-minority" harassment (Burson & Godfrey, 2018; Richeson & Craig, 2011), few studies have directly compared how third-party observers react to majority-on-minority and minority-on-minority acts of hate speech to see how they differ.

Expectations Regarding Minority-on-Minority Racial Hate. Failure to examine how the source's (i.e., offender's) race may affect third-party observers' judgments of hate speech is particularly problematic, given that past expectancy violation research has shown that observers' evaluations of White and minority targets can vary significantly in contexts aside from racial prejudice. For instance, in the context of job hiring, White observers evaluate Black and White (mock) job candidates differently, even when they were presented as having the same qualifications. Jussim et al. (1987) found that skilled Black candidates' positive expectancy violations produced stronger positive evaluations compared to skilled White candidates; on the other hand, Black unskilled candidates were judged more negatively than unskilled White candidates. These evaluations suggest that among White observers tasked with judging the competency and job skills of targets from different racial groups, the effect of expectancy violations was much stronger for Black than White targets.

Relatedly, White observers may hold existing expectations regarding the likelihood and appropriateness of minority-on-minority racial hate speech; however different theoretical perspectives offer contrasting predictions about what they might look like. Observers who hold perspectives motivated by the *common ingroup identity model* (Gaertner et al., 2000) might expect other members of nontargeted minority groups (e.g., Black, Hispanic) to empathize with the surge in pandemic-related online hate speech directed at fellow minority Asian American groups. Those who believe in this "shared fate" model of minority relations should therefore expect individuals from other minority groups to refrain from tweeting racial hate messages against other minorities. This expectation would likely lead observers to judge any kind of racial hate tweet by a (nontargeted) minority source as ethnically nonprototypical and more unexpected compared to a hate tweet posted by a White source.

Alternatively, observers who hold an *intergroup competition perspective* characterize interminority relations as a battle for resources and status within the dominant culture. Thus (in an attempt to deflect the majority group's aspersions and/or jockey for position in larger racial hierarchies, e.g., Kim, 1999) hate tweets from members of other, nontargeted minority groups toward Asian Americans might be viewed as expected, or ethnically prototypical behavior. As the competition for resources like jobs, health care access, and housing has only intensified during COVID-19, those who hold this view might come to expect *more* conflict between minority groups, and thus view anti-Asian tweets posted by a minority source as more normative, rather than nonprototypical.

To summarize, most past studies have explored the situational factors that affect observers' judgments of "prototypical" or "expected" forms of majority-on-minority hate speech, in which a White source attacks a minority target. But because few studies have examined examples of minority-on-minority racial hate speech, we know very little about how observers' expectations and judgments of hate speech by majority and minority sources compare. Existing literature offers two different explanations: First, it is possible that among observers who hold common ingroup identity ideas, instances of minority-on-minority hate speech would be viewed as unorthodox. Thus, a minority source posting anti-Asian tweets would be judged as ethnically nonprototypical, compared to the more expected behavior of a White source. Contrastingly, observers who hold ingroup competition views may find interminority conflict (and related acts of hate speech) to be expected, thus judging anti-Asian tweets from minority sources as ethnically prototypical behavior.

Although both theoretical explanations are plausible, we hypothesize that a White source tweeting COVID-19 anti-Asian hate speech will be evaluated as more ethnically prototypical by outside White observers, consistent with results from studies of "classic" majority-on-minority racial harassment:

> *H1.* Among White observers viewing anti-Asian hate tweets, White sources will be judged as more ethnically prototypical compared to Black sources.

The Effect of Political Partisanship

Another factor that may also affect observers' judgments is *political partisanship*. In the U.S., recent evidence has indicated that the partisan gap on issues of race and racial inequalities has continued to widen. Survey data from many public polling institutes indicate that generally, Democrats are more likely to embrace ideas of racial equality than Republicans. The Public Religion Research Institute (Najle & Jones, 2019) reported that the majority of Democrats in the U.S. welcome the idea of a more racially diverse national population (65%), whereas only 29% of Republicans like the idea of an ethnically diverse country. Academic researchers note similar partisan views about the country's changes in racial demographics, with Republicans reporting higher levels of anxiety about increasing diversity and the decline of the White majority in the U.S. population compared to Democrats (Myers & Levy, 2018).

Yet according to survey data from Pew Research, "Democrats are twice as likely as Republicans to say it has become more common for people to express racist or racially insensitive views since Trump was elected . . . These partisan differences remain when looking only at White Democrats and Republicans" (Horowitz et al., 2019). At least among White adults in the U.S., expectations about the expression of racially insensitive views are strongly associated with political partisanship, such that Democrats may be, overall, more likely to expect to see anti-Asian COVID tweets expressed on Twitter than Republicans. However, we anticipate that Democrats' heightened expectations for anti-Asian hate speech are not equally applicable to both (prototypical) White sources *and* (nonprototypical) minority sources.

Specifically, we predict that political partisanship differences will affect how observers view majority and minority sources of hate speech. We anticipate that as a group, Democrats that embrace racial diversity are more likely to hold common ingroup perspectives on interminority relations. Thus, compared to Republicans, while Democrats may expect that the public expression of racially insensitive views is generally increasing, they would be less likely to expect such views to be expressed by members of other minority groups. Furthermore, we expect that minority-on-minority acts of racial hate speech may be viewed by Democrats as less prototypical or expected, compared to Republicans who are more likely to hold intergroup competition views on minority relations. Overall, then, we predict that there will be interaction effects of *source race* and *observer political partisanship* on observers' judgments of source ethnic prototypicality:

> *H2.* White observers' political partisanship (Democrat/Republican) interacts with source race (Black/White) to affect judgments of message source's ethnic prototypicality such that (a) White Democrat observers will judge White sources tweeting anti-Asian hate speech as more ethnically prototypical compared to Black sources, while (b) White Republican observers will judge Black sources tweeting anti-Asian hate speech as more ethnically prototypical compared to White sources.

Effects of Expectations and Message Evaluations on Subsequent Activism and Allyship

Following EVT, we can hypothesize generally that when observers' expectancies are violated, they engage in greater scrutiny and more careful evaluations of a source and their behavior. As a result of this hypothesized linkage between expectancy violations and related judgments, it is worthwhile considering what kinds of evaluative and behavioral outcomes matter among observers. In this study, we examine three outcomes for observers: (a) judgments of message (tweet) offensiveness, (b) behavioral intention for online activism through signing a petition, and (c) actual activism behavior of clicking on a link to access the online petition. We anticipate that the extremity of observers' evaluation of an anti-Asian tweet as an offensive speech act, and any ensuing behavioral intentions and behaviors, are linked to their initial judgments of the source's prototypicality (i.e., expectancy).

By focusing specifically on those factors that motivate White observers' activism response, we can better understand their inclination to function as allies in the fight against online racism. Allies are "people who recognize the unearned privilege they receive from society's patterns of injustice and take action to change it" (Williams & Sharif, 2021, p. 2). Notably, the Internet has been cited by researchers as a place where allies can easily engage in social justice through online activism, build meaningful online ties with those whom they wish to support, and learn more about communities they hope to help empower. Thus, when a genuine activism response is triggered, an ally can create real social change in online spaces.

However, researchers also point out that allies tend to engage in empty allyship online through "hashtag activism" whereby people profess support (but rarely follow through) with true self-reflection, genuine attitude change, or behavioral response. This kind of "cosmetic" allyship is common in social media and is used "as a means to an end" to manage one's own impression, build a brand, or maintain credibility rather than take action to change inequalities in the status quo (Wellman, 2022). Often "there is an element of performance at play when it comes to ally identity" (Bourke, 2020, p. 185), which refers to the notion of those individuals who only want to adopt the ally label as part of their character and so only portray (rather than genuinely enact) cooperative behavior that truly offers solidarity and support to others (see also, Case, 2012).

Although online activism behaviors have been characterized as low-cost, low-risk "slacktivism," actions such as signing online petitions, or changing one's profile picture in support of a social justice issue have been shown to be strongly correlated with more tangible forms of civic engagement, such as public protest marches or monetary donations (Lee & Hsieh, 2013). Such evidence suggests that online activism can be an important precursor to real attitude change and offline behavior. Furthermore, as the pandemic has limited physical gatherings, online activism becomes a useful proxy to understand how racial hate speech might motivate observers' future (offline) activism and allyship efforts.

In this study, we explore observers' activism behavior in response to hate tweets as their willingness to support the "Stop COVID-19 Disinformation; Stop Anti-Asian Violence" online petition organized by 18 Million Rising (18MR, n.d.), a group that "connects the power of Asian America to digital first organizing." The purpose of this petition was to "tell Facebook, Twitter, and YouTube to immediately shut down hate and misinformation about COVID-19 on their platforms." While we know that behavioral intentions are strongly correlated with actual behaviors (Ajzen & Fishbein, 1974; Rains et al., 2018), we include both a measure of behavioral intentions, as well as an immediate behavioral indicator in which observers were invited to click on a hyperlink that routed them to the 18MR online petition where they could add their signature. In doing so, we could assess whether expectancy judgments about a hate tweet and its source could motivate outside White observers to take any sort of action on behalf of a targeted minority outgroup.

We anticipate an inverse relationship, such that the more prototypical (expected) observers judge the source's behavior to be, the less extreme they will judge the tweet itself, and the less motivated they will be to take any sort of subsequent action. Conversely, when a source's behavior is viewed as unexpectedly nonprototypical, the tweet will be scrutinized more closely and its effects judged as more extreme—or, in this case more offensive. Simultaneously, we expect that viewing unexpected (bad) behaviors performed by others might motivate observers' behavioral intentions to engage in online activism and actual online activism behavior.

> *H3.* White observers' judgments of the source's ethnic prototypicality are (a) negatively related to judgments of tweet offensiveness, (b) negatively related to intentions to engage in online activism, and (c) negatively related to online activism behavior (i.e., clicking on a weblink to an online petition denouncing anti-Asian online disinformation and violence).

Linking these effects together, we hypothesize a moderated-mediation effect, in which the source's race combines with observers' political partisanship to affect observers' judgments of source prototypicality. Observers' source prototypicality judgments are then predicted to affect their message evaluations, online activism behavioral intentions, and activism behavior as follows:

> *H4.* White observers' political partisanship (Democrat/Republican) will moderate the effect of source race (White/Black) on observers' ethnic prototypicality judgments, which will in turn, affect observers' (a) judgments of tweet offensiveness, (b) intentions to engage in online activism, and (c) online activism behavior.

Finally, evidence suggests that we can anticipate relationships between observers' political partisanship and their overall judgments of hate tweets and online activism behaviors. General patterns suggest that political conservatives find online hate content less offensive and disturbing than political liberals do (Costello et al., 2019). More related to the COVID-19 context, being Republican and holding conservative attitudes seems directly associated with anti-Asian attitudes during the pandemic. Holt et al.'s (2022) experimental findings indicated that individuals' political affiliation and party had a direct effect on their perceptions of media framing of the coronavirus. They examined how Democrats/liberals differed from Republicans/conservatives in their judgments of a news article that referenced the coronavirus as either the "Chinese virus" or the "COVID-19 virus." Their results indicated that Democrats perceived the use of "Chinese virus" more negatively than the COVID-19 virus, while Republicans saw no difference between the two articles. Their results also indicated that Republicans held stronger anti-Asian attitudes compared to Democrats and were more likely to blame China for the pandemic. Based on this recent evidence, we expect direct effects of political partisanship on outside observers' judgments and actions:

> *H5.* Compared to White Republican observers, White Democrat observers will (a) judge all hate tweets as more offensive, (b) hold stronger intentions to engage in online activism, and (c) show stronger evidence of online activism behavior.

Method

Sample and Procedure

A sample of 196 White participants ($n_{\text{male}}=95$) who lived in the United States and were over 18 ($M=40.35$, $SD=11.95$) was recruited from TurkPrime (Litman et al., 2017). Participants were routed to an online consent form on the Qualtrics platform where they indicated consent by clicking through to the main survey. On average, participants spent 8 min on the survey ($SD=4.71$ min) and were compensated $2. Participants indicated they had an active Twitter account and used it at least once per week, with a majority of the sample using it daily ($n=131$).[1] After meeting selection criteria, participants indicated their *political partisanship* on both dichotomous (Democrat, $n=122$/Republican, $n=74$) and continuous measures, 1 = "a strong Republican" to 7 = "a strong Democrat" ($M=4.82$, $SD=2.21$).[2]

Experimental Stimuli. The stimuli depicted an anti-Asian tweet being communicated by a male source. Although the source's sex was held constant, the *source's race* systematically varied as either White (ingroup/majority) or Black (outgroup/nontargeted minority). Following past studies (e.g., Munger, 2017) source race was induced through avatar forms in the Twitter profile. Anti-Asian hate tweets during COVID-19 have been wide-ranging; however, because observers' perceptions of message offensiveness or harm often depends on the explicitness of the language being used (Siegel, 2020), we developed stimuli tweets that used the "ChinkFlu" slur seen frequently on Twitter in March 2020. Such an explicit slur was more likely to be perceived outright as racist hate speech by outside observers compared to more ambiguous language. Stimuli were pretested with an offset group of participants prior to use in the main study.[3]

A manipulation check examining the perceived realism of the stimuli (i.e., the likelihood of this kind of content showing up on Twitter) across four items on a 1 = "not realistic" to 7 = "very realistic" scale. A one-sample *t*-test indicated that the tweet realism ratings were significantly above the perceived midpoint of the scale, $t\,(195)=19.28$, $p<.001$, with an average rating of 5.17 ($SD=1.21$).

Measures

Participants responded to four items adapted from Hoffmann et al. (2020) that measured perceptions of the source's *ethnic prototypicality* (e.g., "Based on this tweet, I think the Poster . . . is very similar to others in his ethnic group"; *alpha* = .84, $M=4.94$, $SD=1.20$), with higher scores reflecting stronger judgments of ethnic prototypicality. They also offered their judgments of the tweet's offensiveness across three items (e.g., "In your opinion, how serious is the offense, if any, in this tweet?"; *alpha* = .93, $M=5.35$, $SD=1.65$), and reported their intentions to support the 18MR petition by contributing to the 25,000-signature goal on a 1 = "very unlikely" to 7 = "very likely" scale ($M=4.17$, $SD=2.26$). In addition to intention, we recorded whether participants clicked on the 18MR hyperlink provided at the end of the online survey as an indicator of activism behavior ($n=34$ clicked).

Results

All analyses were conducted using SPSS and the PROCESS macro (Hayes, 2021), using 95% bias-corrected bootstrap confidence intervals based on 10,000 resamples. See Table 1 for a zero-order correlation matrix and descriptive statistics.

Hypothesis 1 predicted that observers judge White sources of anti-Asian hate tweets to be more ethnically prototypical compared to Black sources. An independent samples *t*-test indicated that there was not a significant difference in how participants rated the ethnic prototypicality of Black sources ($M=3.05$, $SD=1.13$) and White sources ($M=3.08$, $SD=1.27$), $t\ (194)=.186$, $p=.852$, Cohen's $d=.03$, rejecting H1.

Hypothesis two predicted that observers' political partisanship moderates (cleaved moderation; see Holbert & Park, 2020) the effect of the tweet source on ethnic prototypicality such that (a) White Democrat observers will judge White sources of anti-Asian hate tweets as more ethnically prototypical compared to Black sources, while (b) White Republican observers will judge Black sources of anti-Asian hate tweets as more ethnically prototypical compared to White sources. Model 1 in the PROCESS macro

Table 1. Descriptive Statistics and Correlations.

Variable	*n*	*M*	*SD*	1	2	3	4
1. Political partisanship	196	4.82	2.21				
2. Source prototypicality	196	3.06	1.20	−.10			
3. Tweet offensiveness	196	5.35	1.65	.53**	−.35**		
4. Online petition intent	196	4.17	2.26	.57**	−.18*	.60**	

$*p<.01$, $**p<.001$.

Table 2. Hypothesis 4 Test of Moderated Mediation for Outcome of Tweet Offensiveness.

	Source Prototypicality (M)				
	Coeff	*SE*	*p*	LLCI	ULCI
Constant	2.71	.27	<.001	2.17	3.25
Tweet Condition (X)	1.28	.40	.002	0.49	2.07
Political Partisanship (W)	.08	.05	.132	−0.24	0.19
Interaction (X*W)	−.27	.08	.001	−0.42	−0.12
	R2 = .07				
	F(3, 192) = 4.87, *p* = .003				
	Tweet Offensiveness (Y1)				
	Coeff	*SE*	*p*	LLCI	ULCI
Constant	6.71	.33	<.001	6.07	7.35
Tweet Condition	.22	.22	.332	−0.22	0.65
Source Prototypicality	−.48	.09	<.001	−0.66	−0.30
	R2 = .13				
	F(2, 193) = 13.94, *p* < .001				
	Conditional Indirect Effects				
Political Partisanship	Coeff	*SE*		LLCI	ULCI
W = 1 (Strong Republican)	−.36	.16		−0.70	−0.08
W = 6	.16	.09		−0.22	0.34
W = 7 (Strong Democrat)	.29	.11		0.07	0.52

n = 196. For Tweet Condition (X), 0 = White, 1 = Black.

was used to test this prediction. The formal test of moderation was significant F (1,192) = 12.56, $p<.001$, $\Delta R^2=.06$. As predicted, strong Republicans found the Black source to be more prototypical, $b=0.74$ ($SE=0.27$), 95% CI = [0.21, 1.27] and strong Democrats found the White source to be more prototypical, $b=-0.60$ ($SE=0.35$), 95% CI = [−1.06, −0.13], supporting H2.

Hypothesis three predicted that White observers' judgments of the source's ethnic prototypicality will be (a) negatively related to judgments of tweet offensiveness, (b) negatively related to intentions to engage in online activism, and (c) negatively related to online activism behavior. Ethnic prototypicality perceptions were negatively associated with tweet offensiveness, $r=-.45$, $p<.001$, and unrelated to behavioral intentions to engage in activism, $r=-.15$, $p=.087$, or actual engagement in online activism, $r_{pb}=-.043$, $p=.627$, providing partial support for H3.

Hypothesis four predicted that White observers' political partisanship will moderate the effect of source race on observers' ethnic prototypicality judgments, which will in turn, affect observers' (a) judgments of tweet offensiveness (Y_1), (b) intentions to engage in online activism (Y_2), and (c) actual online activism behaviors, or clicking on the weblink to the 18MR online petition (Y_3). Model 7 in PROCESS tested the predicted moderated-mediation patterns; estimates of the indirect of effect of the tweet source (X) on each outcome (Y_1, Y_2, Y_3), through ethnic prototypicality judgments (M) were obtained at three levels of political partisanship (W, strong Republicans; political moderates; strong Democrats). For the outcome of tweet offensiveness (Y_1), moderated-mediation was detected, index = .13 ($SE=0.04$), 95% CI = [0.05, 0.22]. A significant indirect effect was detected for strong Republicans, $b=-0.36$ ($SE=0.16$), 95% CI = [−0.70, −0.08] that was in the opposite direction of the significant indirect effect for strong Democrats, $b=0.29$ ($SE=0.11$), 95% CI = [0.07, 0.52].

For the outcome of intent to engage in online activism (Y_2), moderated-mediation was also detected, index = .09 ($SE=0.04$), 95% CI = [0.01, 0.18]. A significant indirect effect was detected for strong Republicans, $b=-0.25$ ($SE=0.14$), 95% CI = [−0.57, −0.02] that was in the opposite direction of the significant indirect effect for strong Democrats, $b=0.20$ ($SE=0.10$), 95% CI = [0.02, 0.42]. There was no evidence of moderated-mediation for actual engagement in online activism (Y_3), index = .04 ($SE=0.04$), 95% CI = [−0.03, 0.14].

Interpreting these relationships indicates that strong Republican observers were more inclined to view the Black source as ethnically prototypical; their expectations regarding the source's behavior were associated with less extreme judgments of tweet offensiveness and less intention to support the 18MR online petition. On the other hand, strong Democrat observers were more inclined to view the White source as being more ethnically prototypical, and these judgments were in turn linked to decreased tweet offensiveness judgments and decreased intention to support the online petition. We return to these results in the discussion section. See Tables 2, 3, and 4 for complete moderated-mediation results.

Hypothesis five stated that compared to White Republican observers, White Democrat observers will (a) judge all hate tweets as more offensive, (b) hold stronger intentions to engage in online activism, and (c) be more likely to engage in online activism by clicking on a link to an online petition. Correlation analyses indicated that political partisanship (1 = strong Republican to 7 = strong Democrat) was significantly correlated with tweet offensiveness, $r=.53$, $p<.001$, and intent to engage in activism, $r=.57$, $p<.001$. A logistic

Table 3. Hypothesis 4 Test of Moderated Mediation for Outcome of Online Petition Intent.

	Online Petition Intent (Y2)				
	Coeff	*SE*	*p*	LLCI	ULCI
Constant	5.13	.47	<.001	4.20	6.06
Tweet condition	.12	.32	.703	−0.51	0.75
Source prototypicality	−.33	.13	.013	−0.60	−0.07
	R2 = .03				
	F(2, 193) = 3.18, *p* = .043				
	Conditional Indirect				
Political partisanship	Coeff	*SE*		LLCI	ULCI
W = 1 (strong republican)	−.25	.14		−0.57	−0.02
W = 6	.11	.07		−0.02	0.27
W = 7 (strong democrat)	.20	.10		0.02	0.42

n = 196. For Tweet Condition (*X*), 0 = White, 1 = Black.

Table 4. Hypothesis 4 Test of Moderated Mediation for Outcome of Online Petition Clicked.

	*Online petition clicked (Y3)				
	Coeff	*SE*	*p*	LLCI	ULCI
Constant	−1.20	.56	.033	−2.29	−0.10
Tweet condition	.21	.38	.580	−0.54	0.96
Source prototypicality	−.16	.17	.335	−0.49	0.17
	Conditional indirect effects				
Political partisanship	Coeff	*SE*	LLCI	ULCI	
W = 1 (strong republican)	−.12	.13	−0.41	0.10	
W = 6	.05	.07	−0.04	0.22	
W = 7 (strong democrat)	.10	.11	−0.07	0.35	

**n* = 196. Logistic regression used for binary outcome (0 = link not clicked; 1 = link clicked).

Table 5. Political Party by Online Petition Clicked.

		Political party		Total
		Democrat	Republican	
Clicked petition link	No	94	68	162
	Yes	28	6	35
Total		122	74	196
Pearson chi-square	Value	df		*p*
	7.08	1		.008

regression indicated that political partisanship significantly predicted actual online activism, χ^2 (1, N=196)=5.37, p=.021. Overall, H5 was supported (see Table 5 for crosstab results).

Discussion

Against the background of the COVID-19 pandemic in the U.S., we explored how White observers passively viewed and evaluated majority-on-minority and minority-on-minority acts of online racial hate speech. We found that overall, anti-Asian tweets were generally perceived as an offensive speech act by all observers; however, the extremity of those tweet perceptions was dependent on evaluations of the tweet's source, which varied as a function of both observers' own political partisanship and the source's race. Results revealed evidence of a moderated-mediation pathway, in which strong Democrats and strong Republicans differed in their judgments of the ethnic prototypicality of White and Black sources of anti-Asian hate tweets. These source prototypicality judgments were in turn associated with observers' judgments of tweet offensiveness and self-reported intentions to engage in online activism.

First, we note that our results revealed the power of expectancies. We found that the more nonprototypical—or unexpected—a source seemed to the observer, the more extreme subsequent judgments of speech act offensiveness and intention to engage in online activism were as well. This is consistent with EVT's logic, which suggests that expectancy violations trigger greater scrutiny of the source of the unexpected behavior and the behavior itself. Specifically, we found that White Republican observers were more likely to judge the minority Black source as more ethnically prototypical than the majority White source, while White Democrat observers judged the White source as being more prototypical and expected than the Black source. These patterns suggest that as third-party observers, White individuals do hold expectations about online racial hate speech: Republicans may not find cross-minority antagonism on Twitter to be that unusual, while White Democrats are more likely to expect other White individuals to use racist language to target minorities. These contrasts in expectancy effects are notable and cover new theoretical ground, as few prior racial hate speech studies have examined these variables in such a vivid context. However, this experiment offers a modest first step into this arena. In the future, researchers should conduct more nuanced examination of people's expectations regarding interminority relations among larger swaths of majority-observer populations to explore how other observer-based factors might affect expectancies, source judgments, and reactions to different kinds of hate speech language.

Although we did find significant effects for judgments of tweet offensiveness and intentions to engage in online activism, the moderated-mediation pattern was not significant for actual online activism behavior. In assessing actual click-through behavior, we found roughly 17% (n=33) of participants in this sample actually clicked on the 18MR petition link. The majority of these clicks were performed by Democrats (80%), but most participants did not click on the link. The point biserial correlation between stated intent to support the petition and clicking the link was $r_{pb}=.35$, $p<.001$, and weaker for Democrats ($n=122$), $r_{pb}=.24$, $p<.001$, than Republicans ($n=74$), $r_{pb}=.46$, $p<.001$. Although Democrats did objectively click through to sign the petition at a higher rate, generally, their reported intentions did not strongly predict their action. Meta-analyses indicate variability in these associations across studies, but they also indicate that some of the most robust associations in communication science have been found between intentions and behavior (e.g., $r=.82$).

Moving forward, future research might examine how features unique to this context—garnering tangible support for marginalized individuals from nontargeted allies—might generate atypically low associations between people's stated intentions and actions. Kalina (2020) uses the term *performative allyship* to refer to individuals from nonmarginalized groups who profess support for a marginalized group but do so to reap personal benefits; this type of allyship can be counterproductive. Wellman (2022) documents examples of performative allyship in the wake of the George Floyd murder, where influencers engaged in hashtag activism to enhance their own credibility rather than to genuinely support the Black Lives Matter movement. Although there might have been practical reasons why participants who stated their intention to support the 18MR campaign did not click on the link in this study (e.g., privacy concerns, lack of trust), future research should seek to better understand not only the factors that might increase individuals stated propensity to support marginalized groups, but also how that sentiment can be most effectively translated into genuine action.

Future work might also examine how characteristics of online platforms influence the type of content people post and observers' reactions. Social media users develop perceptions about what is normative for particular platforms (e.g., Twitter, Instagram, YouTube) and often shape their online behavior to fit normative expectations so as to avoid social sanctions (McLaughlin & Vitak, 2012; Waterloo et al., 2018). From an expectancy violations perspective, it might not only be the source and message that viewers consider when interpreting acts of online hate speech, but also the platform where such speech acts are performed and whether such posts are more or less anticipated. As seen in this study, greater expectancy violations might instantiate stronger rebuke and greater tangible support for those facing discriminatory behavior online.

As with all studies, our work is not without limitations. First, despite pilot testing different, organic tweets for our stimuli, we only used one message in our stimuli to represent the experimental conditions. Stimulus sampling would be beneficial for future studies to better ensure that findings are not due to idiosyncrasies of the specific message and avatars used here (for further reading on single-message designs see Jackson et al., 1989; Slater, 1991). Related to viewing only one tweet, although all participants found the tweet realistic, they still might have viewed it as anomalous to what typically appears on Twitter given it was only one tweet from one source. If participants saw multiple messages from different sources, they may have been more inclined to sign the online petition.

We also note that our experiment involved specific combinations of majority-on-minority (White/Asian) and minority-on-minority (Black/Asian) online racial hate speech during COVID-19. Although the context of this study was driven by the pandemic, we recognize that these design choices may reflect the ongoing trope of Black-Asian conflict often found in the mass media (Lee & Huang, 2021; Yang, 2021). However, if we look carefully, these results also speak to the complexities underlying racial stratification in the U.S. In considering the larger White-Black-Asian dynamic, Kim's (1999, 2022); *racial triangulation theory* argues that Asian Americans occupy unique positions along two dimensions relative to Blacks and Whites: The *valorization* dimension reflects rankings of "superiority/inferiority" with Whites at the top, Blacks at the bottom, and Asians in a middle position that is more advantageous compared to Blacks, but lower than Whites. Asian Americans' middle position on the valorization axis is reflected in the "model minority" stereotype that lauds Asians for their work ethic, intelligence, and relative success. However, on the *civic ostracism* dimension that ranks groups on insider-ness/foreignness, Kim (1999) asserts that Asians occupy the lowest place: Their "forever foreigner" status cements them as being unable to assimilate into a (White-dominated) American culture. As a result, Asians assume an even lower position than Blacks on this axis.

Scholars note the dialectical relationship of these two dimensions and how they play out in these seemingly opposite stereotypes; yet even today, many people still rely on them to make sense of the Asian American experience. For example, racial othering of Asian Americans has become even more prominent during the pandemic through use of the forever foreigner stereotype that paints Asian Americans as "dishonest, diseased invaders" who brought the coronavirus into the U.S. (Li & Nicholson, 2021, p. 4). In addition, Whites have also been shown to endorse racial valorization of Asian Americans to a greater extent than Black observers (Xu & Lee, 2013).

It is possible that endorsement of model minority and forever foreigner stereotypes may not only shape White observers' views about Asian Americans, but also how they expect members of other minority groups to interact with them. Kim (2022) points out how the racial valorization of Asians can disavow the disadvantages experienced by Black Americans by pitting two minority groups against each other in an intergroup competition fashion. Relatedly, triangulation dynamics may also affect White individuals' views on their own status: Given their dominant position atop both racial hierarchical dimensions, some White observers in this study may have anticipated greater allyship among members of minority groups—in other words, expecting greater Asian-Black unification in the fight for racial equality. As unrealistic and unfair as they are, expectations are often set in which "Black people are framed as the necessary caretakers of racial minorities beyond themselves" (Chen & Hosam, 2022, p. 456). Keeping this in mind, it becomes clear how some observers could judge anti-Asian hate tweets from a Black source as a much greater violation of expectancies than that same tweet from a (prototypical) White source.

However, because we did not test them directly, these kinds of alternative expectations remain speculative. As ours is a preliminary study, we hope researchers continue to examine how different minority groups—both targeted and nontargeted—process and respond to acts of online hate speech; there currently exists a paucity of research on interminority relations in the context of online discrimination. We note that these experimental stimuli (though carefully designed and pretested) were not intended to be representative of the myriad of contexts, experiences, and characteristics of online racial hate speech. Instead, our initial experiment offers important, but preliminary insights into the phenomenon of cross-minority racism. Testing the effects of expectancies across different forms and contexts of online discrimination would be pertinent to establish their influence over the attitudinal and behavioral reactions of third-party viewers who routinely observe and evaluate acts of racial hate in their social media feeds.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Communication Association.

ORCID iD

Stephanie Tom Tong https://orcid.org/0000-0002-1612-4425

Notes

1. A post hoc power analysis using G* Power indicated that our sample was well powered to detect a medium size effect ($f = .25$; power = .94) but underpowered for the detection of small effects ($f = .10$; power = .29).
2. Below, although we only report results with the continuous variable, replications using the dichotomous variable ("Democrat"/"Republican") were conducted and results were similar to those reported in the text.
3. We pretested several versions of tweets containing more/less explicit language to use in our experimental stimuli. Results of our pretest indicated that a tweet containing an explicit racial slur of "ChinkFlu" ($M = 5.33$) was rated as more offensive than those containing more ambiguous racially centered phrases of "KungFlu" ($M = 4.33$) or "ChinaVirus" ($M = 4.16$), $F(2, 58) = 5.35$, $p = .007$. Following these results, we incorporated "ChinkFlu" into the experimental stimuli for use in the main studies.

References

18 Million Rising (18MR). (n.d.). *Stop COVID-19 disinformation. Stop anti-Asian violence*. https://action.18mr.org/stop-antiasian-violence/

Ajzen, I., & Fishbein, M. (1974). Factors influencing intentions and the intention-behavior relation. *Human Relations*, *27*(1), 1–15. https://doi.org/10.1177/001872677402700101

Barnidge, M., Kim, B., Sherrill, L. A., Luknar, Ž., & Zhang, J. (2019). Perceived exposure to and avoidance of hate speech in various communication settings. *Telematics and Informatics*, *44*, 101263. https://doi.org/10.1016/j.tele.2019.101263

Baron, R. S., Burgess, M. L., & Kao, C. F. (1991). Detecting and labeling prejudice: Do female perpetrators go undetected? *Personality and Social Psychology Bulletin*, *17*(2), 115–123. https://doi.org/10.1177/0146167291017002001

Bettencourt, B. A., Manning, M., Molix, L., Schlegel, R., Eidelman, S., & Biernat, M. (2016). Explaining extremity in evaluation of group members: Meta-analytic tests of three theories. *Personality and Social Psychology Review*, *20*(1), 49–74. https://doi.org/10.1177/1088868315574461

Bliuc, A. M., Faulkner, N., Jakubowicz, A., & McGarty, C. (2018). Online networks of racial hate: A systematic review of 10 years of research on cyber-racism. *Computers in Human Behavior*, *87*, 75–86. https://doi.org/10.1016/j.chb.2018.05.026

Bourke, B. (2020). Leaving behind the rhetoric of allyship. *Whiteness and Education*, *5*(2), 179–194. https://doi.org/10.1080/23793406.2020.1839786

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communications Monographs*, *55*(1), 58–79. https://doi.org/10.1080/03637758809376158

Burgoon, J. K., & Walther, J. B. (1990). Nonverbal expectancies and the evaluative consequences of violations. *Human Communication Research*, *17*(2), 232–265. https://doi.org/10.1111/j.1468-2958.1990.tb00232.x

Burson, E., & Godfrey, E. B. (2018). The state of the union: Contemporary interminority attitudes in the United States. *Basic and Applied Social Psychology*, *40*(6), 396–413. https://doi.org/10.1080/01973533.2018.1520106

Case, K. A. (2012). Discovering the privilege of whiteness: white women's reflections on anti-racist identity and ally behavior. *Journal of Social Issues*, *68*, 78–96. https://doi.org/10.1111/j.1540-4560.2011.01737.x

Chen, S. G., & Hosam, C. (2022). Claire Jean Kim's racial triangulation at 20: rethinking Black Asian solidarity and political science. *Politics, Groups, and Identities*, *10*(3), 455–460. https://doi.org/10.1080/21565503.2022.2044870

Costello, M., Hawdon, J., Bernatzky, C., & Mendes, K. (2019). Social group identity and perceptions of online hate. *Sociological Inquiry*, *89*(3), 427–452. https://doi.org/10.1111/soin.12274

Cowan, G., & Hodge, C. (1996). Judgments of hate speech: The effects of target group, publicness, and behavioral responses of the target. *Journal of Applied Social Psychology*, *26*(4), 355–374. https://doi.org/10.1111/j.1559-1816.1996.tb01854.x

Cowan, G., & Mettrick, J. (2002). The effects of target variables and setting on perceptions of hate speech. *Journal of Applied Social Psychology*, *32*(2), 277–299. https://doi.org/10.1111/j.1559-1816.2002.tb00213.x

Daniels, J. (2017, October 19). Twitter and White supremacy: A love story. *DAME Magazine*. https://www.damemagazine.com/2017/10/19/twitter-and-white-supremacy-love-story/

DeTurk, S. (2011). Allies in action: The communicative experiences of people who challenge social injustice on behalf of others. *Communication Quarterly*, *59*(5), 569–590. https://doi.org/10.1080/01463373.2011.614209

Gaertner, S. L., Dovidio, J. F., Nier, J. A., Banker, B. S., Ward, C. M., Houlette, M., & Loux, S. (2000). The Common Ingroup Identity Model for reducing intergroup bias: Progress and challenges. In D. Capozza & R. Brown (Eds.), *Social identity processes: Trends in theory and research* (pp. 133-148). SAGE.

Gilbert, D. (2020, March 27). Anti-Chinese hate speech online has skyrocketed since the Coronavirus crisis began. *Vice*. https://www.vice.com/en_us/article/n7jywd/anti-chinese-hate-speech-online-has-skyrocketed-since-the-coronavirus-crisis-began

Hawdon, J., Oksanen, A., & Räsänen, P. (2017). Exposure to online hate in four nations: A cross-national consideration. *Deviant Behavior*, *38*(3), 254–266. https://doi.org/10.1080/01639625.2016.1196985

Hayes, A. (2021). PROCESS macro.: https://www.processmacro.org/index.html

Hoffmann, P., Platow, M. J., Read, E., Mansfield, T., Carron-Arthur, B., & Stanton, M. (2020). Perceived self-in-group prototypicality enhances the benefits of social identification for psychological well-being. *Group Dynamics: Theory, Research, and Practice*, *24*(4), 213–226. https://doi.org/10.1037/gdn0000119

Holbert, R. L., & Park, E. (2020). Conceptualizing, organizing, and positing moderation in communication research. *Communication Theory*, *30*(3), 227–246. https://doi.org/10.1093/ct/qtz006

Holt, L. F., Kjærvik, S. L., & Bushman, B. J. (2022). Harm and shaming through naming: Examining why calling the coronavirus the "COVID-19 Virus," not the "Chinese Virus," matters. *Media Psychology*, *25*, 639–652. https://doi.org/10.1080/15213269.2022.2034021

Horowitz, J. M., Brown, A., & Cox, K. (2019, April 9). Race in American 2019. *Pew Research Center*. https://www.pewresearch.org/social-trends/2019/04/09/views-of-racial-inequality/

Jackson, S., O'Keefe, D. J., Jacobs, S., & Brashers, D. E. (1989). Messages as replications: Toward a message-centered design strategy. *Communication Monographs*, *56*(4), 364–384. https://doi.org/10.1080/03637758909390270

Jussim, L., Coleman, L. M., & Lerch, L. (1987). The nature of stereotypes: A comparison and integration of three theories. *Journal of Personality and Social Psychology*, *52*(3), 536–546. https://doi.org/10.1037/0022-3514.52.3.536

Kalina, P. (2020). Performative allyship. *Technium Social Sciences Journal*, *11*, 478–481.

Kenski, K., Coe, K., & Rains, S. (2020). Perceptions of uncivil discourse online: An examination of types and predictors. *Communication Research*, *47*, 795–814. https://doi.org/10.1177/0093650217699933

Kim, C. J. (1999). The racial triangulation of Asian Americans. *Politics & Society*, *27*(1), 105–138. https://doi.org/10.1177/0032329299027001005

Kim, C. J. (2022). Asian Americans and Anti-Blackness. *Politics, Groups, and Identities*, *10*(3), 503–510. https://doi.org/10.1080/21565503.2021.2016448

Lee, J., & Huang, T. (2021, March 11). Why the trope of Black-Asian conflict in the face of anti-Asian violence dismisses solidarity. *Columbia University*.: https://sociology.columbia.edu/news/why-trope-black-asian-conflict-face-anti-asian-violence-dismisses-solidarity

Lee, Y. H., & Hsieh, G. (2013, April). *Does Slacktivism hurt activism? The effects of moral balancing and consistency in online activism*. In Proceedings of the SIGCHI conference on human factors in computing systems. ACM Digital Library. https://doi.org/10.1145/2470654.2470770

Leets, L. (1999, May). *A cultural perspective on racist speech harm*. [Paper presentation]. 49th International Communication Association, San Francisco, CA, United States.

Leets, L. (2001). Explaining perceptions of racist speech. *Communication Research*, *28*(5), 676–706. https://doi.org/10.1177/009365001028005005

Leets, L. (2003). Disentangling perceptions of subtle racist speech: A cultural perspective. *Journal of Language and Social Psychology*, *22*(2), 145–168. https://doi.org/10.1177/0261927X03022002001

Leets, L., & Giles, H. (1997). Words as weapons—when do they wound? Investigations of harmful speech. *Human Communication Research*, *24*(2), 260–301. https://doi.org/10.1111/j.1468-2958.1997.tb00415.x

Li, Y., & Nicholson, H. L., Jr. (2021). When "model minorities" become "yellow peril"—Othering and the racialization of Asian Americans in the COVID-19 pandemic. *Sociology Compass*, *15*(2), e12849. https://doi.org/10.1111/soc4.12849

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z

Marti, M. W., Bobier, D. M., & Baron, R. S. (2000). Right before our eyes: The failure to recognize non-prototypical forms of prejudice. *Group Processes & Intergroup Relations*, *3*(4), 403–418. https://doi.org/10.1177/1368430200003004005

Mathew, B., Kumar, N., Goyal, P., & Mukherjee, A. (2018). Analyzing the hate and counter speech accounts on Twitter. https://arxiv.org/pdf/1812.02712.pdf#:~:text=After%20the%20annotation%20process%2C%20the,hatespeech%20spread%20by%20the%20user

McLaughlin, C., & Vitak, J. (2012). Norm evolution and violation on Facebook. *New Media & Society*, *14*(2), 299–315. https://doi.org/10.1177/1461444811412712

Meyers, C., Leon, A., & Williams, A. (2020). Aggressive confrontation shapes perceptions and attitudes toward racist content online. *Group Processes & Intergroup Relations*, *23*(6), 845–862. https://doi.org/10.1177/1368430220935974

Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5

Myers, D., & Levy, M. (2018). Racial population projections and reactions to alternative news accounts of growing diversity. *The Annals of the American Academy of Political and Social Science*, *677*(1), 215–228. https://doi.org/10.1177/0002716218766294

Najle, M., & Jones, R. P. (2019). American democracy in crisis: The fate of pluralism in a divided nation. *Public Religion Research Institute*. https://www.prri.org/research/american-democracy-in-crisis-the-fate-of-pluralism-in-a-divided-nation/

Rains, S. A., Levine, T. R., & Weber, R. (2018). Sixty years of quantitative communication research summarized: Lessons from 149 meta-analyses. *Annals of the International Communication Association*, *42*, 105–124. https://doi.org/10.1080/23808985.2018.1446350

Richeson, J. A., & Craig, M. A. (2011). Intra-minority inter-group relations in the Twenty-First century. *Daedalus*, *140*, 166–175. https://doi.org/10.1162/daed_a_00085

Searle, J. (1965). What is a speech act? In M. Black (Ed.), *Philosophy in America* (pp. 221-239). Cornell University Press.

Sellars, A. F. (2016). *Defining hate speech*. Berkman Klein Center Publication. http://dx.doi.org/10.2139/ssrn.2882244

Siegel, A. A. (2020). Online hate speech. In N. Persily & J. A. Tucker (Eds.). *Social media and democracy*: *The state of the field, prospects for reform* (pp. 56–88). Cambridge University Press.

Slater, M. D. (1991). Use of message stimuli in mass communication experiments: A methodological assessment and discussion. *Journalism Quarterly*, *68*(3), 412–421. https://doi.org/10.1177/107769909106800312

Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, *44*(2), 136–146. https://doi.org/10.1002/ab.21737

Tong, S. T., Stoycheff, E., & Mitra, R. (2022). Racism and resilience of pandemic proportions: Online harassment of Asian Americans during COVID-19. *Journal of Applied Communication Research*, *50*(6), 595–612. https://doi.org/10.1080/00909882.2022.2141068

Tong, S. T., & Walther, J. B. (2015). The confirmation and disconfirmation of expectancies in computer-mediated communication. *Communication Research*, *42*(2), 186–212. https://doi.org/10.1177/0093650212466257

Walther, J. B. (in press). Online hate: A prosocial explanation of antisocial behavior and affordances of social media. In R. Nabi & J. Myrick (Eds.), *Our online emotional selves: The link between digital media and emotional experience*. Oxford University Press.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, *20*(5), 1813–1831. https://doi.org/10.1177/1461444817707349

Wellman, M. L. (2022). Black squares for Black lives? Performative allyship as credibility maintenance for social media influencers on Instagram. *Social Media + Society*, *8*, 20563051221080473. https://doi.org/10.1177/20563051221080473

Williams, M., & Sharif, N. (2021). Racial allyship: Novel measurement and new insights. *New Ideas in Psychology*, *62*, 100 865.

Xu, J., & Lee, J. C. (2013). The marginalized “model” minority: An empirical examination of the racial triangulation of Asian Americans. *Social Forces*, *91*(4), 1363–1397. https://doi.org/10.1093/sf/sot049

Yang, J. (2021, May 17). The “Black-Asian Conflict” is a problematic trope—and it’s time to end it. *Medium*. https://stopasianhate.medium.com/the-black-asian-conflict-is-a-problematic-trope-and-it-s-time-to-end-it-cef2d4176099

Author Biographies

Stephanie Tom Tong (PhD, Michigan State University) is an associate professor in the Communication Department at Wayne State University. She is the Director of the Social Media and Relational Technologies (SMART) Labs, which investigates how technology affects the ways people communicate across a variety of relational contexts including romance, families, friendships, personal health, and online hate.

David C. DeAndrea (PhD, Michigan State University) is an associate professor and the Director of Graduate Studies in the School of Communication at the Ohio State University. His research examines how features of communication technology affect the way people evaluate information and strategically manage impressions online.