ACR

ASIAN COMMUNICATION RESEARCH
September 2021, Vol. 18, No. 2, pp. 51-68
e-ISSN 2765-3390, p-ISSN 1738-2084
https://doi.org/10.20879/acr.2021.18.2.51

Original Research

The Effects of Racial Hate Tweets on Perceived Political Polarization and the Roles of Negative Emotions and Individuation

Jayeon Lee [1]

[1] Department of Media & Communication, Gachon University, Republic of Korea

Corresponding author
Jayeon Lee
Department of Media & Communication, Gachon University, 1342 Seongnam-daero, Bokjeong-dong, Sujeong-gu, Seongnam-si, Gyeonggi-do 13120, Republic of Korea
Email: lee.jayeon@gmail.com

Disclose statement
No potential conflict of interest to disclose was reported by the author.

Received
17 May 2021

Revised
15 July 2021

Accepted
10 August 2021

ABSTRACT

While there are growing concerns among Americans about widening partisan gaps, some claim that the perceptions of polarization are exaggerated. According to the theory of false polarization, people tend to assume that opposing parties are more ideologically extreme than they actually are. Hate speech expressed and disseminated through social media, which provokes anger among the audience, might be a factor that drives such heightened perceptions. Drawing on the social identity model of deindividuation effects (SIDE model) and the affect-cognition literature, this experimental study (N = 480) investigates whether exposure to hate messages on Twitter (i.e., hate tweets) targeting Whites triggers negative emotions (i.e., anger and fear), thereby increasing the perceptions of political polarization among Whites. This study also examines whether the effect of hate tweets decreases when the source is individuated with personal information disclosed on the profile. A conditional process analysis reveals that exposure to hate tweets significantly provokes negative emotions and heightens perceived polarization through anger, and that source individuation significantly moderates the anger-provoking effect of hate tweets. These findings indicate that online hate speech may have implications for political perception and anger can play a significant role in the process. The theoretical and practical implications are discussed.

KEYWORDS

polarization, hate speech, social media, anger, individuation

Polarization is one of the keywords that define current American society. A growing body of research demonstrates that political divisions within the public have deepened considerably since the 1970s (Abramowitz & Saunders, 1998; Jacobson, 2004). During the 2016 U.S. presidential campaign, partisans' views of the opposing party were reported to be more negative than at any point in the past 22 years (Pew Research Center, 2016). Some claim that American politics has reached peak polarization (Drutman, 2016) although others argue that the public's actual issue positions have not changed as much as people

think (Fiorina & Abrams, 2008; Levendusky & Malhotra, 2016; Mason, 2013). Whether the perception is accurate or not, perceived political polarization has implications for individuals and society, including citizens' understanding of issues, general trust, and political discussion and participation (Abramowitz & Saunders, 1998; Gervais, 2015; McCright & Dunlap, 2011). Thus, it is important to understand what might lead individuals to form such perceptions.

Hate speech generated and disseminated online can be a reason why we find more divisions among people. Research has shown that polarization is substantially influenced by a citizen's broader political environment, such as social media platforms (Klein & Robison, 2019). Scholars point out that there is a need to understand the operation of prejudicial discourses in non-media contexts such as online discussion boards (Chovanec, 2021). Interactions on Twitter are known to aggravate opinion polarization and inter-group hostility (Yarchi et al., 2020). While hateful messages and inflammatory images on Twitter (i.e., hate tweets) have been a serious social problem, the microblogging site has been criticized for failing to actively deal with hate speech (Baden, 2017; Meyer, 2016). A study found that people who saw uncivil online comments perceived larger gaps between Republicans and Democrats than those who did not see them (Hwang et al., 2014). However, the mechanism by which uncivil content affects perceived gaps has not been fully explored.

The role negative emotions play is one of the main foci of this study. Compared to other areas of study such as psychology, the functions of emotions have remained theoretically and empirically understudied in communication (Himelboim et al., 2016). While it is known that social media use often triggers negative emotions, and such emotions can amplify individuals' perceptions or cognitions (Chovanec, 2021; Yarchi et al., 2020), few attempts have been made to investigate the role of negative emotions as mediators in the processes of political cognitions. The present study aims to fill this gap in the literature and examine how so-called hate tweets (i.e., hate speech on Twitter) affect the target group members' (i.e., Whites) perceptions of political polarization by provoking negative emotions.

Drawing on the social identity model of deindividuation effects (SIDE model) and the affect-cognition literature, the present study predicts that exposure to hate tweets (i.e., hate speech on Twitter) triggers negative emotions, and in turn, increases the perceptions of social or political polarization. This study examines whether negative emotions play the role of mediators between hate tweet exposure and perceived polarization. In addition, it investigates the role of source individuation as a potential moderator. That is, this study examines whether the effect of hate tweets on perceived polarization increases or decreases depending on the amount of personal information about the hate speaker. Based on the test of the final moderated mediation model, the implications of the results for understanding the theoretical linkages between hate speech on social media and perceived political polarization are discussed.

Perceptions of Polarization

It is widely believed that American society has become increasingly divided and polarized over key political or social issues. Pew Research Center's 2017 report indicates that the average gap between Republicans and Republican-leaning independents and Democrats and Democratic-leaning independents across 10 issues (e.g., the environment, homosexuality, immigration, and economic policy) increased from 15 percent points in 1994 to 36 percent points in 2017 and to 39 percent points in 2019 (Doherty, 2017; Pew Research Center, 2019b). According to the recent Pew report (2019a), the

majority of Americans think that there are major differences in what the parties stand for; they think that both are too extreme in their positions and that Republicans and Democrats "not only disagree over plans and policies, but also cannot agree on the basic facts." Partisans say that they do not share their political as well as nonpolitical values and goals with those of the opposing party. Half the Democrats say the Republican Party makes them "afraid" (55%) and "angry" (47%), while half the Republicans say the same about the Democratic Party (49% and 46%, respectively) (Pew Research Center, 2016).

Are these negative perceptions of opposing parties accurately based on their actual issue stances? Perceived political polarization here means the extent to which an individual believes politics to be divided along ideological or party lines—mainly between two major parties in the U.S. (Ploger, 2019) and it can be different from actual political polarization. According to the theory of false polarization (Pronin et al., 2002; Robinson et al., 1995; Sherman et al., 2003), people tend to assume that opposing parties are ideologically consistent and extreme. The pro-party rhetoric used by party elites to appeal to in-group (vs. out-group) bias and the partisan media that amplify the rhetoric also lead to polarized perceptions, particularly among partisans (Fiorina et al., 2005; Levendusky & Malhotra, 2016; Lu & Lee, 2019; Yang et al., 2016).

As a result, although it is believable that Democrats and Republicans hold more distinctive ideological positions and policy attitudes than in the past, the partisan gap tends to be overestimated. Partisans' biased perceptions of other parties' positions were found in various issue contexts, including abortion (Robinson et al., 1995) and social welfare (Farwell & Weiner, 2000). Exposure to partisan conflict-framed news was found to polarize news consumers along party lines (Han & Federico, 2017). Mason (2015) argues that the current perception of ideological polarization has been heightened by in-group bias, activism, and anger increase, which are influenced more by partisan sorting (i.e., holding identities that are aligned with their party) than by actual opinion extremity.

Social Identity Model of Deindividuation Effects (SIDE) Model

While many social and emotional factors can drive perceptions of polarization, it is important for researchers to investigate potential moderators of the process. One such variable this study focuses on is individuation. From the standpoint of the social identity model of deindividuation effects (SIDE model), polarization may be understood as a result of group identification-seeing people simply as part of the group they belong to, rather than unique individuals with idiosyncratic characteristics (Lee, 2007; Postmes & Spears, 1998; Spears et al., 2001). Group identification becomes heightened when little individuating information is provided, which makes each person's group memberships more salient (Maslach et al., 1985; Spears et al., 1990). This situation results in stronger adherence to group norms and greater opinion polarization.

The effects of group identification on perceptions can be larger in a computer-mediated communication setting where individuating cues are less available than in face-to-face settings. One study found that perceived polarization was significantly and consistently related to online news consumption in nine of the ten countries examined, suggesting the possibility that the online information people typically consume promotes biased perceptions of others (Yang et al., 2016). Group-based perceptions are also likely to occur in social media environments because messages are shared and filtered by like-minded people in users' social networks, and members of the opposing groups can be easily mocked or portrayed in biased and extreme ways within the networks (Conover et al., 2011).

Group conflicts also lead to heightened

perceptions of differences between categories through an increased salience of group identity (Spears et al., 2001; Tajfel, 1981). Political conflicts and extreme voices among political elites are often highlighted by the media, thereby distorting people's perception of polarization (Han & Federico, 2017; Hwang et al., 2014; Yang et al., 2016). Incivility, which can signal group conflicts to observers, is of particular relevance to online communication and its impact on polarization. Incivility such as vitriolic words and name-calling, is prevalent in online posts and comments and has political implications such as decreasing political trust and efficacy (Borah, 2013; Mutz & Reeves, 2005). In Hwang et al.'s (2014) study, incivility significantly widened the perceived political gaps between Republicans and Democrats. It is noteworthy that they found a significant association between incivility and perceived polarization, not polarization itself. These findings lead us to predict that situations where people notice group conflicts somehow widen perceived gaps between groups. In a similar vein, racial hate speech is also likely to increase perceptions of polarization, whether it matches the actual level of polarization between parties.

Hate Speech and Anti-White Tweets

According to the United Nations (2019), hate speech is "any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor" (p. 2). Heyman (2018) sees hate speech as the most extreme forms of speech, such as incivility and vilification. According to him, incivility comprises name calling, derision of the opponent, and impoliteness, while vilification is calling the other side evil, malevolent, and dangerous. He claims that hate speech goes one step beyond vilification, and dehumanizes others, on the basis of their race, ethnicity, religion, gender, or sexual orientation as though they are not entitled to respect. Online hate speech, or cyberhate, is a growing phenomenon (Kaakinen et al., 2018; Sood et al., 2012). Being correlated with actual hate crimes, hate speech can pose threats and have serious consequences for its victims (Williams et al., 2020). However, while many scholars have investigated the existence and emergence of hate speech on social media (see Bliuc et al., 2018), relatively little is known about its influence on social media users.

Most existing research on hate speech focuses on racial/ethnic minority groups such as African Americans. This is natural given that racism is conceptualized as discrimination or antagonism directed against people on the basis of their minority or marginalized racial or ethnic group membership (Wetherell & Potter, 1992). Anti-African American bias is the most common factor (48.6%) that motivates racial hate crime, and Whites are the most commonly known offenders (50.7%) of racial hate crime (FBI, 2017). It has been known that the discourse of hate increased in the U.S. since Donald Trump whose presidency was characterized as his impoliteness and deeply offensive remarks toward immigrants and ethnic minorities (Wodak et al., 2020). The second most common cause of racial hate crime is anti-White bias (17.1%). This is notable because Whites are the historic and current majority of people living in the U.S., constituting 72% of the population (United States Census Bureau, 2010).

Twitter is a good platform for investigating social conflicts. Among social media outlets, Twitter has been known as a political platform that facilitates polarization and hostility compared to Facebook or WhatsApp (Yarchi et al., 2020). Twitter has another distinct characteristic with regard to its user base. Since its launch in 2006, Twitter has been popular among racial minorities. According to a 2021 report by the Pew Research Center, 29% of

Black Internet users and 23% of Hispanic Internet users were on Twitter, outpacing non-Hispanic Whites (22%). Twitter has been a major social media platform for Black activism. According to Twitter's announcement on its 10th anniversary, the most-used Twitter social issue hashtag is #Ferguson, and #BlackLivesMatter is the third most used (Anderson, 2016). In 2017, #TakeAKnee, which was inspired by San Francisco 49ers quarterback Colin Kaepernick's protest against police brutality of Blacks, received more than 56,000 retweets and 180,000 likes in three days. Black Americans' active use of Twitter to form tight clusters and support the Black community even gave birth to a new term: "Black Twitter."

The presence of the strong Black community on Twitter has prompted anti-Black sentiments among some White Twitter users. Silva and colleagues (2016) analyzed hate tweets in the sampled tweets between June 2014 and June 2015, and found that the most frequently targeted group was Blacks; "Nigga" (31.11%; #1 target word), "Black people" (4.91%; #4), and "Nigger" (1.84%; #9) were included in the top 10 list of target words, accounting for almost 40% of the total. The second most frequent target word was "White people" (9.76%). Other targets on the list were not race-specific (e.g., stupid people). It is observed that the conflict between Whites and non-Whites has risen since Donald Trump's campaign (Rhodes et al., 2017). An Asian American *New York Times* writer's old tweets mocking White people prompted angry reactions from them, igniting a debate on the concept of "anti-White racism" (Rosenberg & Logan, 2018). Considering the ongoing conflicts among racial groups, it is worth investigating how racial minorities' anti-majority messages on Twitter could affect the majority. Thus, this study focuses on the impact of hate speech toward Whites (Caucasians). It is not the intention of this study to advocate the concept of "anti-White racism" or "reverse racism" by any means. To begin with, racism requires the power to oppress a group of people based on a social construction, which makes racism from minority groups toward dominant ones impossible (Wetherell & Potter, 1992). Historically and socially, anti-white sentiments are not comparable to anti-black sentiments. Having said that, whether there is a potential influence of anti-White speech on Whites' perceptions of political polarization is a different story and a scholarly question that has been rarely asked.

Negative Emotions: Anger and Fear

An important goal of the present study is to investigate why exposure to online hate speech results in exacerbated perception of polarization. Emotion is a likely factor that underlies the influence of uncivil online messages. Uncivil political discourse breeds negative emotions such as anger and fear (Gervais, 2015; Wang & Silva, 2018). Anger is the most common emotional reaction to uncivil behavior in everyday life (Phillips & Smith, 2004), and it is the predominant response to social discrimination (Matheson & Anisman, 2012). Message creators sometimes intentionally use outrage as the major tool to maximize diffusion of inflammatory content on social media and even trigger violent or self-destructive actions (Vasterman, 2018). Gervais (2015) found that uncivil discourse targeting one's in-group generates anger among the public, while like-minded incivility does not. Negative reactions become stronger when the uncivil attack is aimed at the self (Kinney & Segrin, 1998). Therefore, hate speech on social media that unjustifiably attacks a racial group is likely to trigger anger among the target group members. Thus, the following hypothesis is posited:

H1a: Exposure to hate tweets will increase the target group members' anger.

Along with anger, fear seems to be relevant to

such hate speech situations. Like anger, fear is a negative emotion that tends to be experienced after mistreatment by others (Phillips & Smith, 2004). Symbols of hatred trigger fear in those who are targeted. For example, racial supremacy/extremist groups such as the Ku Klux Klan frequently use symbols like swastikas, and these intimidate those who are being hated, triggering fear and insecurity among them, while giving haters a sense of power and belonging (Anti-Defamation League, 2016). Similarly, hate speech is known to create "an atmosphere of fear" (Lederer & Delgado, 1995, p. 5). Hate tweets against specific racial groups are likely to elicit fear among the targeted members. Thus, the following hypothesis is posited:

H1b: Exposure to hate tweets will increase the target group members' fear.

Mediating Role of Negative Emotions

Spinoza once defined emotions as "states that make the mind inclined to think one thing rather than another" (Frijda et al., 2000, p. 1). Affect-cognition literature has demonstrated the power of mood and emotion to profoundly influence cognitive evaluations and judgments of objects or persons (Clore et al., 1994; Frijda et al., 2000; Schwarz & Clore, 1996). In this research tradition, affect is used as a generic term to refer to both moods and emotions (Forgas, 1995). Incidental affect triggered by emotional experiences often influences judgments and choices that are irrelevant to the experiences. When positive feelings (e.g., happiness or joy) or negative feelings (e.g., sadness, anger, or fear) were induced by having participants recall their memories in a specific domain, their perceived self-efficacy was affected not only regarding activities in the same domain, but also in other remote areas (Kavanaugh & Bower, 1985; Schwarz & Clore, 1983). For example, when people were reminded of their failure in a past romantic relationship, their estimates of success in academic or athletic activities were also reduced below the neutral level. This phenomenon is called "affect infusion" and defined as "the process whereby affectively loaded information exerts an influence on and becomes incorporated into the judgmental process, entering into the judge's deliberations and eventually coloring the judgmental outcome" (Forgas, 1995, p. 39). Affect infusion is believed to occur because emotional states make mood-congruent thoughts more available, thereby having a widespread impact on perceptions and judgments.

Notably, moods and emotions influence not only people's judgments about themselves but also the perceptions of others or society. In Johnson and Tversky's (1983) study, participants induced to feel negative affect made more pessimistic judgments or estimates of given problems (e.g., frequencies of death) than participants induced to feel positive affect. In one of the first affect-cognition studies, Razran (1940) found that sociopolitical messages were deemed significantly more positive when the audience was happy (after receiving free food) rather than unhappy (after being exposed to unpleasant odors). In another classic study, participants who were made fearful through electric shocks tended to perceive another person as fearful and anxious (Feshbach & Singer, 1957). Many other studies in the affect infusion framework found that those in negative affect made more negative estimates and judgments than those in positive affect (Clore et al., 1994; Schwarz & Clore, 1996).

In political communication, negative emotions have been found to play mediating roles between the use of pro-party media and affective polarization; exposure to partisan information provoke viewers' negative emotions, resulting in more extreme feelings toward themselves or opposing parties and candidates (Lu & Lee, 2019). It might be not only affective polarization, but also perceived polarization that the media influence through negative emotions. Anger in

particular is often associated with intergroup conflict and aggression against the out-group (Mackie et al., 2000). In a study, anger and Republican identification interacted in predicting perceived polarization (Huber et al., 2015). According to Huber et al. (2015), intergroup conflict can drive social cognitive tendencies, such as perceived polarization, by making people angry. This suggests that negative emotions may function a mediator between a stimulus and its perceptual outcome. This mediating role of negative emotion can also explain why uncivil online discourse leads viewers to perceive more polarization (Hwang et al., 2014). It also guides us to predict that anger and fear play a mediating role between racial hate speech and perception of polarization. Thus, the following set of hypotheses is posited:

H2: (a) Anger and (b) fear will mediate the relationship between exposure to racial hate tweets and perceived polarization such that exposure to racial hate tweets will increase perceived polarization by provoking anger among the target group members.

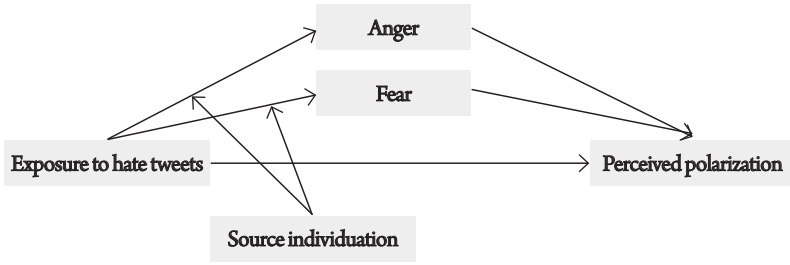
Moderating Role of Individuating Information

Group identification becomes heightened when little individuating information is provided. Playing down idiosyncratic characteristics of individuals leads to reduced sense of individual identity and immersion in a group (Maslach et al., 1985; Spears et al., 1990). This situation results in stronger adherence to group norms and greater opinion polarization. Prior research compared the results of the high and low amount of personal information about communication partners, and found that people were more likely to draw in-group vs. out-group distinctions and have extreme perceptions of other groups when they lacked individuating cues about others (Lee, 2006, 2007; Postmes et al., 2002; Spears et al., 2001; Tajfel, 1981).

In a similar vein, members of groups attacked by racial hate tweets are more likely to be emotionally agitated (i.e., angry and fearful) when they feel that the hate speaker is a representative of their whole group. When the source is seen as an unknown individual, it is easier to ignore them. Thus, while racial hate speech sparks negative emotions, and, in turn, increases the perceptions of political polarization, the absence of individuating information on one's social media profile is likely to heighten the effects (see Figure 1). Thus, the following hypotheses are posited:

H3: Source individuation will moderate the effects of hate tweets on perceived polarization through decreases in (a) anger and (b) fear.

Figure 1. *The Hypothesized Model*



METHOD

Participants

Participants for the experiment were recruited through the Qualtrics online panel management team between October 28 and November 3, 2016 ($N = 480$), right before the U.S. presidential election. Among Qualtrics online panel members, only adult U.S. citizens who self-identified as White were sampled (51.3% female; age $M = 47.34$, $SD = 16.54$). The average education level was between college and two-year college degree. The average income was approximately $50,000.

Study Design

An online experiment site was created for this study. Those who satisfied the recruitment criteria (i.e., age and racial background) were invited to participate through a recruitment email circulated by the Qualtrics online panel management team. To minimize possible experimental demand characteristics (Orne, 1962), participants were informed that the goal of the study was to explore Twitter users' political opinions. Upon agreeing to the consent form, participants were randomly assigned to one of the four experimental conditions based on 2 (tweet type: hate vs. non-hate) × 2 (source individuation: high vs. low) and were exposed to a mock Twitter profile. Participants were informed that they would be asked to view a randomly selected Twitter user profile. All four stimuli profiles were ostensibly owned by an individual whose ID was @JZW0001. In the control (non-hate-tweet) condition ($n = 238$), as a baseline, five tweets that did not express any negative sentiment against a particular racial/ethnic group (e.g., "Happy #NationalDrinkBeerDay!" and "I don't understand why people put spinach on a pizza") were presented. In the hate-tweet condition ($n = 242$), there were three additional tweets that included derogatory and hateful statements against Whites (that is, "Came across 6 inches away from running over a white man with my car, kinda wish I would have," "The only white man you can trust is a dead white man," and "So sick of this. There are NO-GOOD white people. There are only LESS BAD white people.") in addition to the five common control tweets. These three racist messages were adapted from existing hate tweets and were rephrased to fit the purpose and context of the present study.

The manipulation of source individuation was similar to the method used in Lee (2006, 2007), where the amount of individuating information was varied by disclosing/not disclosing personal profile information (e.g., age, hobbies, and favorite TV shows). The high individuation condition ($n = 229$) presented the user profile with information revealing personal identity: "Son, brother, husband, father, dog lover, UFC fan, admirer of Jackie Chan, Disney hater, searcher of great food & cool music." For the source's profile photo, an old movie poster was posted mainly to avoid revealing his own race. The wide banner at the top of the Twitter profile showed the walking legs of a man in jeans and a dog following him. The man's face or skin color was not shown in the image. In the low individuation condition ($n = 251$), there was no personal description, banner, or a unique profile photo.

After seeing one of the four stimuli, participants filled out a questionnaire that included their perceptions of polarization and questions measuring negative emotions, party identification, and demographic variables as well as two manipulation check items. Upon completion, they were debriefed and thanked.

Measures

Negative Emotions (Anger and Fear)

After seeing the Twitter stimulus, participants rated their feeling states on a 7-point Likert type scale (1 = *not at all*; 7 = *very much*). Based on the

Positive and Negative Affect Schedule-Expanded Form (PANAS-X; Watson & Clark, 1994), "hostile," "angry," and "upset" were combined for anger (Cronbach's α = .92; M = 2.70, SD = 1.97), and "afraid," "frightened," and "fearful" (Cronbach's α = .97; M = 2.15, SD = 1.73) were used for fear.

Perceived Political Polarization

The measurement of perceived polarization was adapted from a single-item measure from the 2016 Pew Research Center survey ("how different are the policy positions of the Republican and Democratic parties?"). Participants rated the extent to which they thought the Republican Party and the Democratic Party differed from each other regarding major social/political issues our society was currently facing. A 5-point scale from "not at all different" to "extremely different" was used for the question (M = 3.73, SD = 0.97).

Control Variables

The perceived gap between the two major parties might be a function of partisanship. A study found that the effect of anger on perceptions of polarization differed depending on which party people identify with (Huber et al., 2015). Thus, participants' major party identifications were controlled. Self-identified Republicans were dummy-coded as Republican (n = 138), while self-identified Democrats were dummy-coded as Democrat (n = 140). Those who did not support either of the two major parties (n = 202) were the reference group.

RESULTS

Manipulation Check

Two true or false questions were included to identify if participants recognized the experimental manipulations (i.e., the presence of hate tweets or personal information on the profile). A chi-square test of independence indicated that the difference was statistically significant between the hate and non-hate conditions in the way they reported about the profile, $\chi^2(1, N = 480) = 146.80$, $p < .001$. Specifically, 88% of those who viewed the hate tweets correctly remembered the presence of racially offensive content. In addition, 73% of those who viewed the individuated profile correctly remembered that the profile owner's personal favorites were displayed on the profile, $\chi^2(1, N = 480) = 73.27$, $p < .001$.

Hypotheses Tests

The basic mediation model of this study predicted that exposure to hate tweets would induce anger (H1a) and fear (H1b) and that these would mediate the effect of hate tweets on polarized perception (H2a–b). The final moderated mediation model predicted that the hypothesized mediation would be further moderated by the presence of individuating information (H3a–b). The PROCESS macro model 4 and model 7 (Hayes, 2013) were used to generate 95% bias-corrected bootstrap confidence intervals (CIs) based on 10,000 bootstrap samples.

Table 1. *Bivariate Correlations with Means and Standard Deviations of Key Variables*

Measure	1	2	3	*M*	*SD*
1. Anger	-	.751**	.103*	2.7	1.97
2. Fear		-	0.056	2.15	1.73
3. Polarization			-	3.73	.97

Note. *$p < .05$. **$p < .01$.

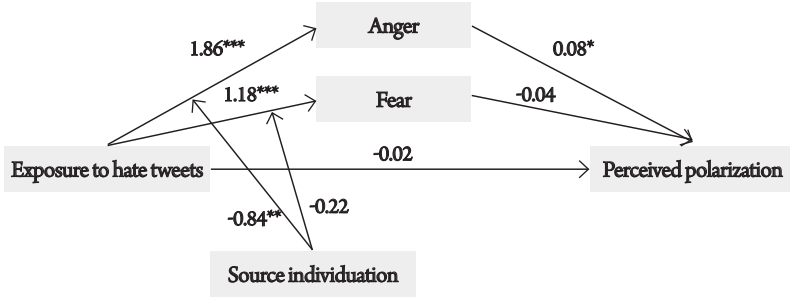
First, the relationship between hate tweets and negative emotions was explored. As expected, those who were exposed to the hate tweets, relative to those who were not, reported greater anger (M = 3.64, SD = 2.08 vs. M = 1.74, SD = 1.29) and fear (M = 2.75, SD = 2.02 vs. M = 1.54, SD = 1.10).

Next, the simple mediation from hate tweet exposure to perceived polarization through negative emotions was examined. Both anger and fear were included in the model as parallel mediators. Hate tweets significantly predicted anger (b = 1.88, SE = 0.16, t = 11.92, p < .001, CI [1.57, 2.19]) and fear (b = 1.19, SE = 0.15, t = 8.03, p < .001, CI [0.90, 1.48]). Thus, both H1a and H1b were supported.

The results of mediation tests showed contrasting outcomes between anger and fear. Anger significantly mediated the effects of hate tweets (b = 0.14, bootstrap SE = 0.07, bootstrap CI [0.02, 0.26]). However, fear was not significant, with the CI including zero (b = -0.05, bootstrap SE = 0.04, bootstrap CI [-0.13, 0.04]). The direct effect of hate tweet exposure on perceived polarization was not significant (b = -0.02, SE = 0.10, t = -0.24, p = .808, CI [-0.22, 0.17]). Thus, only H2a was supported, H2b was not.

Finally, whether the presence of individuating information moderated the indirect effect of hate tweets on perceived polarization through anger (H3a) was investigated. A significant interaction between exposure to hate tweets and source individuation in predicting anger was found (b = -0.84, SE = 0.31, t = -3.06, p < 0.01, CI [-1.45, -0.23]). The indirect effects of hate tweets on perceived polarization through anger significantly differed depending on the individuation condition, making the moderated mediation significant (b = -0.06, bootstrap SE = 0.04, bootstrap CI [-0.15, -0.01). Hate tweets provoked much more anger in the deindividuation condition (b = 1.31, t = 8.35, p < .001, CI [1.00, 1.62]) than in the individuation condition (b = 0.64, t = 3.95, p = <.001, CI [0.32, 0.96]). The moderated mediation model through fear was not significant (b = -0.01, bootstrap SE = 0.02, bootstrap CI [-0.02, 0.05]). Thus, only H3a was supported.[1]

Figure 2. *The Final Moderated Mediation Model*



Note. Control variables are not included in the figure.
*p < .05. **p < .01. ***p < .001.

[1] Both Republican partisanship and Democratic partisanship were significant predictors in the final moderated mediation model (b = 0.31, SE = 0.11, t = 2.91, p < .01, CI [0.10, 0.52], and b = 0.40, SE = 0.11, t = 3.78, p < .001, CI [0.19, 0.61], respectively).

DISCUSSION

Incivility in online environments and political polarization have been among the issues that damage democratic communication processes. However, the perceptual aspect of polarization and its relationship to incivility have not been fully investigated. The present study attempted to fill this gap. The findings revealed that those who were exposed to social media messages derogating their in-group and became angry saw deeper polarization along political party lines than those who were not. It was also found that the moderating role of source individuation significantly adjusted the impact of hate tweets.

These findings have several implications. First, it suggests that our current perceptions of political polarization can partially be a function of prevailing incivility that we are exposed to on the Internet and anger it triggers. The increasing volume of uncivil online discourse raises the need for scholars to examine its consequences from various angles. Previous research has found that racial hate tweets negatively influence the psychological and physical well-being of target racial group members (e.g., Lee-Won et al., 2017). Studies in political communication found that online incivility contributed to polarized perceptions about an issue (Anderson et al., 2014). Building on that, the present findings revealed that hate tweets could also influence perceptions of polarization through negative emotion.

It is crucial to extend our understanding of perceived polarization because exaggerated perceptions of political divide erode citizens' belief in and expectations about public deliberation (Hwang et al., 2014). Perception is often more influential than reality in politics, as demonstrated in the political communication literature (Mutz, 1998; Noelle-Neumann, 1984). Political perceptions can be the basis of political opinion and behaviors, and can foster actual polarization in society. For instance, perceived polarization may prompt more partisan sorting, affecting how they position themselves about political issues, and whether they express their opinion or participate in politics. Strong partisans may engage more in selective exposure in their information search when they perceive wide gaps between major parties, while weak or nonpartisans may engage in partisan sorting, become stronger partisans, or shy away from engaging in politics (Lupu, 2015; Yang et al., 2016).

Second, the findings of the present study have important implications regarding the role of negative emotion in political perception. Specifically, the results of the mediation test provide empirical evidence that anger can be a vehicle that lets hate tweets influence our perception of polarization. It is notable that reading strongly uncivil tweets targeting Whites did not significantly affect White viewers' perceptions if they did not feel enough anger. While the role of emotion has been largely overlooked in studies of political polarization, the mediation model supported by the current data demonstrates the importance of viewers' emotional reactions to daily experiences in determining its impact on their judgment. This finding is in line with the affect infusion literature that highlights the vital role of emotions in cognition and judgment. According to Forgas (1995), affect infusion is not likely to occur when the target judgment is easy, well-known, or personally relevant. Given that affect infusion occurred in the present study, it seems that polarization is something people feel, but find it difficult to quantify clearly.

Meanwhile, it should be noted that the mediating role of fear was not significant. While hate tweets provoked fear as well as anger, feeling fear after reading the tweets did not significantly affect perception of polarization. It is not clear what caused the difference; however, anger and joy are in general experienced more frequently and exert stronger effects than

sadness and fear (Lerner et al., 2003; Scherer, 1986). If participants experienced anger as the dominant emotional reaction, their level of fear might not be enough to have an independent effect on perceptions of polarization. Affect intelligence theory may also offer some clues. According to it, anxiety leads people to be more attentive and deliberative, which motivates systematic information processing, while anger facilitates heuristic processing (Marcus et al., 2000). Considering that both fear and anxiety are responses to a known (fear) or unknown (anxiety) threat or danger, fear might also facilitate systematic information processing, which keeps one from jumping into hasty affect-based judgments.

Lastly, it is noteworthy that source individuation can moderate the strong impact of hate tweets on anger. By simply revealing a little more personal information through the Twitter profile, even without identity-revealing photos, the worldview-changing effects of uncivil messages could be significantly reduced. This finding is in line with the SIDE model perspective that deindividuation obscures within-group differences, which in turn can induce group polarization (Lee, 2007; Spears et al., 1990; Spears et al., 2001).

Methodologically, by taking an experimental approach, this study enabled the establishment of causal links between online hate speech, target group members' emotional reactions, and political perceptions. The present experiment, conducted exclusively with White participants, demonstrates that exposure to uncivil social media messages attacking a racial/ethnic group can color the group members' perceptions, even when they are the majority group in society and not a typical target of racism. As anti-White messages are increasingly circulated through social media, it is important to recognize that these have implications for Whites' worldviews.

This study had several limitations that need to be addressed. First, an artificial social media setting was used for the experiment. To tease out the manipulation effect while controlling confounders, this study relied on a screenshot image of a Twitter page on which participants could not scroll down or click. Other information that is provided by Twitter, such as the profile owner's interactions with his followers, was not available in the current experimental setting. Participants' reactions to hate speech could be different in a real social media setting.

Another limitation concerns the measurement of the dependent variable, perceived polarization, which was measured using a single item scale. Although quite a few other studies have employed a similar single-item instrument for measuring perceived polarization (e.g., Hwang et al., 2014; Pew Research Center, 2016), a single-item measure might not be ideal to fully represent the concept, particularly when its purpose is to measure general perceptions. This study also allows us to measure a short-term effect only because the perception question was asked shortly after exposing participants to hate tweets. It will be beneficial if future research examines long-term ramifications of anger-provoking social media messages for perceived polarization using multiple items. As the sample was exclusively focused on Whites, future studies that replicate the current findings with other racial group members are strongly encouraged. It will be interesting to examine how seeing these anti-White tweets affects non-Whites' polarization perceptions.

Despite these limitations, the findings of this study are compelling. They contribute to the existing body of literature by shedding light on the understudied area of research: implications of hate speech for political perception and the role of anger. While racial hate tweets have been pointed out as a serious issue, their consequences for users' political perceptions have not been fully investigated. A network analysis of Twitter during the 2010 U.S. congressional midterm found that the network of political retweets exhibits a highly

segregated partisan structure, with extremely limited connectivity between left- and right-leaning users (Conover et al., 2011). Considering that those who have few or no friends in the opposing party are much more likely to have hostile feelings about the other (Pew Research Center, 2016), staying in an echo chamber with only like-minded people can facilitate the "us vs. them" perception. In this environment, a few hateful remarks from unknown others can easily amplify the perceived gaps between groups, which in turn can exacerbate actual political polarization. Thus, a few requirements for disclosing personal information might help reduce the exaggerated perception of polarization. This can be practical implications for online communities.

The findings of the present study offer initial evidence that online hate speech and personal, individuating information of social media user profiles may have broader implications for political perceptions than one might think. Hopefully, the findings will further stimulate research on social media and political perception and encourage communication scholars to further examine the long-term political consequences of online hate speech.

REFERENCES

Abramowitz, A. I., & Saunders, K. L. (1998). Ideological realignment in the U.S. electorate. *The Journal of Politics*, *60*(3), 634-652. https://doi.org/10.2307/2647642

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

Anderson, M. (2016, August 15). 3. *The hashtag #BlackLivesMatter emerges: Social activism on Twitter*. Pew Research Center. https://www.pewresearch.org/internet/2016/08/15/the-hashtag-blacklivesmatter-emerges-social-activism-on-twitter/

Anti-Defamation League (2021). *Hate on display Hate symbols database*. https://www.adl.org/hate-symbols

Baden, R. (2017, August 8). This artist sprayed racist tweets in front of Twitter's German HQ. *VICE*. https://www.vice.com/en/article/7x85eg/this-artist-sprayed-racist-tweets-in-front-of-twitters-german-hq

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

Borah, P. (2013). Interactions of news frames and incivility in the political blogosphere: Examining perceptual outcomes. *Political Communication*, *30*(3), 456-473. https://doi.org/10.1080/10584609.2012.737426

Chovanec, J. (2021). 'Re-educating the Roma? You must be joking...': Racism and prejudice in online discussion forums. *Discourse & Society*, *32*(2), 156-174. https://doi.org/10.1177/0957926520970384

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition*. (2nd ed., Vol.1, pp. 323-417). Psychology Press. https://doi.org/10.4324/9781315807102

Conover, M., Ratkiewicz, J., Francisco, M. R., Gonalves, B., Menczer, F., & Flammini, A. (2011, July 17-21). Political polarization on Twitter [Paper presentation]. *Proceedings of the International AAAI Conference on Web and Social Media*, Barcelona, Catalonia, Spain. https://ojs.aaai.org/index.php/ICWSM/article/view/14126

Doherty, C. (2017, October 5). *Key takeaways on Americans' growing partisan divide over*

political values. Pew Research Center. https://www.pewresearch.org/fact-tank/2017/10/05/takeaways-on-americans-growing-partisan-divide-over-political-values/

Drutman, L. (2016, May 24). *American politics has reached peak polarization*. Vox. https://www.vox.com/polyarchy/2016/3/24/11298808/american-politics-peak-polarization

Farwell, L., & Weiner, B. (2000). Bleeding hearts and the heartless: Popular perceptions of liberal and conservative ideologies. *Personality and Social Psychology Bulletin*, *26*(7), 845-852. https://doi.org/10.1177/0146167200269009

Federal Bureau of Investigation (2018). *Hate Crime Statistics, 2017. U.S. Department of Justice*. https://ucr.fbi.gov/hate-crime/2017

Feshbach, S., & Singer, R. D. (1957). The effects of fear arousal and suppression of fear upon social perception. *The Journal of Abnormal and Social Psychology*, *55*(3), 283-288. https://doi.org/10.1037/h0042811

Fiorina, M. P., & Abrams, S. J. (2008). Political polarization in the American public. *Annual Review of Political Science*, *11*, 563-588. https://doi.org/10.1146/annurev.polisci.11.053106.153836

Fiorina, M. P., Abrams, S. J., & Pope, J. C. (2005). *Culture war? The myth of polarized America* (2nd ed.). Pearson Longman.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *117*(1), 39-66. https://doi.org/10.1037/0033-2909.117.1.39

Frijda, N. H., Manstead, A. S. R., & Bem, S. (2000). *Emotions and beliefs: How feelings influence thoughts*. Cambridge University Press. https://doi.org/10.1017/CBO9780511659904

Gervais, B. T. (2015). Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, *12*(2), 167-185. https://doi.org/10.1080/19331681.2014.997416

Han, J., & Federico, C. M. (2017). Conflict-framed news, self-categorization, and partisan polarization. *Mass Communication and Society*, *20*(4), 455-480. https://www.tandfonline.com/doi/full/10.1080/15205436.2017.1292530

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. The Guilford Press.

Heyman, S. (2018, July 12). *Q+A: How do we solve incivility in society?* The University of Auckland. https://www.thebigq.org/2018/07/12/how-do-we-solve-incivility-in-society/

Himelboim, I., Sweetser, K. D., Tinkham, S. F., Cameron, K., Danelo, M., & West, K. (2016). Valence-based homophily on Twitter: Network analysis of emotions and political talk in the 2012 presidential election. *New Media & Society*, *18*(7), 1382-1400. https://doi.org/10.1177/1461444814555096

Huber, M., Van Boven, L., Park, B. & Pizzi, W. T. (2015). Seeing red: Anger increases how much Republican identification predicts partisan attitudes and perceived polarization. *PLoS One*, *10*(9), e0139193. https://doi.org/10.1371/journal.pone.0139193

Hwang, H., Kim, Y., & Huh, C. U. (2014). Seeing is believing: Effects of uncivil online debate on political polarization and expectations of deliberation. *Journal of Broadcasting & Electronic Media*, *58*(4), 621-633. https://doi.org/10.1080/08838151.2014.966365

Jacobson, G. C. (2004). Explaining the ideological polarization of the congressional parties since the 1970s. *Available at SSRN 1157024*. https://ssrn.com/abstract=1157024

Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, *45*(1), 20-31. https://doi.org/10.1037/0022-3514.45.1.20

Kaakinen, M., Oksanen, A., & Räsänen, P. (2018). Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach. *Computers in Human Behavior*, *78*, 90-97. https://doi.org/10.1016/j.chb.2017.09.022

Kavanagh, D. J., & Bower, G. H. (1985). Mood and self-efficacy: Impact of joy and sadness on perceived capabilities. *Cognitive Therapy and Research*, *9*(5), 507-525. https://doi.org/10.1007/BF01173005

Kinney, T., & Segrin, C. (1998). Cognitive moderators of negative reactions to verbal aggression. *Communication Studies*, *49*(1), 49-72. https://doi.org/10.1080/10510979809368518

Klein, E., & Robison, J. (2020). Like, post, and distrust? How social media use affects trust in government. *Political Communication*, *37*(1), 46-64. https://doi.org/10.1080/10584609.2019.1661891

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*(4), 602-616. https://doi.org/10.1037/0022-3514.79.4.602

Lederer, L., & Delgado, R. (1995). Introduction. In L. Lederer & R. Delgado (Eds.), *The price we pay: The case against racist speech, hate propaganda, and pornography* (pp. 3-13). Hilland Wang.

Lee, E.-J. (2006). When and how does depersonalization increase coformity to group norms in computer-mediated communication? *Communication Research*, *33*(6), 423-447. https://doi.org/10.1177/0093650206293248

Lee, E.-J. (2007). Deindividuation effects on group polarization in computer-mediated communication: The role of group identification, public-self-awareness, and perceived argument quality. *Journal of Communication*, *57*(2), 385-403. https://doi.org/10.1111/j.1460-2466.2007.00348.x

Lee-Won, R. J., Lee, J. Y., Song, H., & Borghetti, L. (2017). "To the bottle I go... to drain my strain": Effects of microblogged racist messages on target group members' intention to drink alcohol. *Communication Research*, *44*(3), 388-415. https://doi.org/10.1177/0093650215607595

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science*, *14*(2), 144-150. https://doi.org/10.1111/1467-9280.01433

Levendusky, M., & Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes? *Political Communication*, *33*(2), 283-301. https://doi.org/10.1080/10584609.2015.1038455

Lu, Y., & Lee, J. K. (2019). Partisan information sources and affective polarization: Panel analysis of the mediating role of anger and fear. *Journalism & Mass Communication Quarterly*, *96*(3), 767-783. https://doi.org/10.1177/1077699018811295

Lupu, N. (2015). Party polarization and mass partisanship: A comparative perspective. *Political Behavior*, *37*(2), 331-356. https://doi.org/10.1007/s11109-014-9279-z

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

Mason, L. (2013). The rise of uncivil agreement: Issue versus behavioral polarization in the American electorate. *American Behavioral Scientist*, *57*(1), 140-159. https://doi.org/10.1177/0002764212463363

Mason, L. (2015). "I disrespectfully agree": The differential effects of partisan sorting on social and issue polarization. *American Journal of Political Science*, *59*(1), 128-145. https://doi.org/10.1111/ajps.12089

Maslach, C., Stapp, J., & Santee, R. T. (1985). Individuation: Conceptual analysis and

assessment. *Journal of Personality and Social Psychology*, *49*(3), 729-738. https://doi.org/10.1037/0022-3514.49.3.729

Matheson, K., & Anisman, H. (2012). Biological and psychosocial responses to discrimination. In J. Jetten, C. Haslam, & A. S. Haslam (Eds.), *The social cure: Identity, health and well-being* (pp. 133-153). Psychology Press.

McCright, A. M., & Dunlap, R. E. (2011). Cool dudes: The denial of climate change among conservative white males in the United States. *Global Environmental Change*, *21*, 1163-1172. https://doi.org/10.1016/j.gloenvcha.2011.06.003

Meyer, R. (2016, July 21). Twitter's famous racist problem. *The Atlantic*. https://www.theatlantic.com/technology/archive/2016/07/twitter-swings-the-mighty-ban-hammer/492209/

Mutz, D. C. (1998). *Impersonal influence: How perceptions of mass collectives affect political attitudes*. Cambridge University Press. https://doi.org/10.1017/CBO9781139175074

Mutz, D. C., & Reeves, B. (2005). The new videomalaise: Effects of televised incivility on political trust. *American Political Science Review*, *99*(1), 1-15. https://doi.org/10.1017/S0003055405051452

Noelle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, *24*(2), 43-51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776-783. https://doi.org/10.1037/h0043424

Pew Research Center. (2014, June 12). *Political polarization in the American public*. https://www.pewresearch.org/politics/2014/06/12/political-polarization-in-the-american-public/

Pew Research Center. (2016, June 22). *Partisanship and political animosity in 2016*. https://www.pewresearch.org/politics/2016/06/22/partisanship-and-political-animosity-in-2016/

Pew Research Center. (2019a, October 10). *Partisan antipathy: More intense, more personal*. https://www.pewresearch.org/politics/2019/10/10/partisan-antipathy-more-intense-more-personal/

Pew Research Center. (2019b, December 17). I*n a politically polarized era, sharp divides in both partisan coalitions*. https://www.pewresearch.org/politics/homepage-image/

Pew Research Center. (2021, April 7). *Social media fact sheet*. https://www.pewresearch.org/internet/fact-sheet/social-media/#who-uses-each-social-media-platform?menuItem=b14b718d-7ab6-46f4-b447-0abd510f4180

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

Ploger, G. (2019, November 26). *Causes and consequences of perceived polarization*. https://doi.org/10.17605/OSF.IO/7UTDR

Postmes, T., & Spears, R. (1998). Deindividuation and anti-normative behavior: A meta-analysis. *Psychological Bulletin*, *123*(3), 238-259. https://doi.org/10.1037/0033-2909.123.3.238

Postmes, T., Spears, R., & Lea, M. (2002). Intergroup differentiation in computer-mediated communication: Effects of depersonalization. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 3-16. https://doi.org/10.1037/1089-2699.6.1.3

Postmes, T., Spears, R., Sakhel, K., & de Groot, D. (2001). Social influence in computer-mediated

communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin*, *27*(10), 1243-1254. https://doi.org/10.1177/01461672012710001

Pronin, E., Puccio, C., & Ross, L. (2002). Understanding misunderstanding: Social psychological perspectives. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 636-665). Cambridge University Press. https://doi.org/10.1017/CBO9780511808098

Razran, G. H. (1940). Conditioned response changes in rating and appraising sociopolitical slogans. *Psychological Bulletin*, *37*, 481-493.

Rhodes, J. H., Schaffner, B. F., & McElwee, S. (2017). Is America more divided by race or class? Race, income, and attitudes among whites, African Americans, and Latinos. *The Forum*, *15*(1), 71-91. https://doi.org/10.1515/for-2017-0005

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "Naive realism" in intergroup perception and conflict. *Journal of Personality and Social Psychology*, *68*(3), 404-417. https://doi.org/10.1037/0022-3514.68.3.404

Rosenberg, E., & Logan, E. B. (2018, August 3). An Asian American woman's tweets ignite a debate: Is it okay to make fun of white people online? *The Washington Post*. https://www.washingtonpost.com/news/the-intersect/wp/2018/08/03/an-asian-american-womans-tweets-ignite-a-debate-is-it-okay-to-make-fun-of-white-people-online/

Scherer, K. R., & Tannenbaum, P. H. (1986). Emotional experiences in everyday life: A survey approach. *Motivation and Emotion*, *10*(4), 295-314. https://doi.org/10.1007/BF00992106

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513-523. https://doi.org/10.1037/0022-3514.45.3.513

Schwarz, N., & Clore, G. L. (1996). Feelings and phenomenal experiences. In E. T. Higgins & A. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433-465). The Guilford Press.

Sherman, D. K., Nelson, L. D., & Ross, L. D. (2003). Naïve realism and affirmative action: Adversaries are more similar than they think. *Basic and Applied Social Psychology*, *25*(4), 275-289. https://doi.org/10.1207/S15324834BASP2504_2

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016, March 17-20). Analyzing the targets of hate in online social media [Poster presentation]. *Proceedings of the International AAAI Conference on Web and Social Media*, Cologne, Germany. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13147

Sood, S., Antin, J., & Churchill, E. (2012, May 5-10). Profanity use in online communities [Paper presentation]. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, Texas, USA.

Spears, R., Lea, M., & Lee, S. (1990). De-individuation and group polarization in computer-mediated communication. *British Journal of Social Psychology*, *29*(2), 121-134. https://doi.org/10.1111/j.2044-8309.1990.tb00893.x

Spears, R., Postmes, T., Lea, M., & Watt, S. E. (2001). A SIDE view of social influence. In J. P. Forgas & K. D. Williams (Eds.), *Social influence: Direct and indirect processes* (pp. 331-350). Psychology Press.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge University Press

United Nations (2019, May). *United Nations strategy and plan of action on hate speech*. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf

United States Census Bureau (2010). *2010 census shows America's diversity*. https://www.census.gov/newsroom/releases/

archives/2010_census/cb11-cn125.html
Vasterman, P. (2018). *From media hype to Twitter storm*. Amsterdam University Press.
Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior, 81*, 73-83. https://doi.org/10.1016/j.chb.2017.11.041.
Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form*. Unpublished manuscript, University of Iowa, Iowa City. https://ir.uiowa.edu/psychology_pubs/11/
Wetherell, M., & Potter, J. (1992). *Mapping the language of racism: Discourse and the legitimation of exploitation*. Columbia University Press.
Williams, M. L., Burnap, P., Javed, A., Liu, H., & Ozalp, S. (2020). Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology, 60*(1), 93-117. https://doi.org/10.1093/bjc/azz049
Wodak, R., Culpeper, J., & Semino, E. (2021). Shameless normalisation of impoliteness: Berlusconi's and Trump's press conferences. *Discourse & Society, 32*(3). https://doi.org/10.1177/0957926520977217
Yang, J., Rojas, H., Wojcieszak, M., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Jones, P. K., Mazzoleni, G., Papathanassopoulos, S., Rhee, J. W., Rowe, D., Soroka, S., Tiffen, R. (2016). Why are "others" so polarized? Perceived political polarization and media use in 10 countries. *Journal of Computer-Mediated Communication, 21*(5), 349-367. https://doi.org/10.1111/jcc4.12166
Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication, 38*(1-2), 98-139. Advance online publication. https://doi.org/10.1080/10584609.2020.1785067