Aggressive confrontation shapes perceptions and attitudes toward racist content online

Chanel Meyers[a1], Angelica Leon[b], & Amanda Williams[c]

York University[a], Department of Psychology
4700 Keele Street, Toronto, ON M3J 1P3, Canada

Teachers College, Columbia University, Department of Psychology[b]
525 West 120th Street, New York, NY, 10027, USA

University of Bristol[c], Graduate School of Education
35 Berkeley Square, Bristol, BS8 1JA, United Kingdom

Please address correspondence regarding this manuscript to the first author.
Email: meyersck@whitman.edu

Accepted at *Group Processes & Intergroup Relations*

[1] The author has since moved to Whitman College, Department of Psychology, 345 Boyer Avenue, Walla Walla, WA, USA.

Abstract

With more people using social media on a daily basis and the prevalence of racial discrimination online, it becomes imperative to understand what factors impact minority individuals' perceptions of these transgressions in an online context. Confrontation to discrimination in the form of comments on social media may meaningfully shape perceptions of racism online. Across three studies, we examine how confrontation type (aggressive vs. passive) and confronter group membership (ingroup vs. outgroup) influence Asian Americans' perceptions of online prejudice and attitudes towards the confronters. In Study 1, we find that aggressive confrontations alter perceptions of a racist online post to be more offensive as compared to passive confrontations. In Study 2, these findings extend to participants' likelihood to report the content as offensive. Lastly, in Study 3, we find that aggressive confronters are evaluated more positively than passive confronters. These findings have important implications for understanding racial discrimination in an online context by demonstrating the impact of confrontation type on minority individuals' perceptions and behaviors.

Keywords: confronting discrimination, racism, group membership, social media

Aggressive confrontation shapes perceptions and attitudes toward racist content online.

The increasing popularity of using social networking online has many benefits, including connecting people and providing platforms for social justice movements, but unfortunately, instances of racial prejudice and discrimination persist on these sites. Ethnic and racial minorities experience racial discrimination – the unfair treatment of or bias towards a person due to their racial or ethnic group membership – across a variety of online domains, including gaming, fandom, news and sport threads and commentaries, social networking sites, and blogs (see Daniels, 2012 for a review). When presented with/encountering online racial discrimination, users have several possible reactions: they may choose to ignore the remark completely, defuse the comment, or boldly confront the racist transgression. One way in which people may see confrontation to racism online is through comments or reactions to social media posts online. This type of confrontation is especially important in an online context, as previous research has found that even the mere presence of online comments can shift attitudes toward current events (Winter et al., 2015), therefore it is possible that online content may also shift perceivers' reactions to racial discrimination.

While descriptive research conducted in online contexts has examined the frequency and forms of discrimination that occur on social networking sites (Cleland, 2014; Jakubowicz et al., 2017; Johns & McCosker, 2015; Pew Research Center, 2016), research has yet to examine how confronting these instances of racial discrimination might shape racial minorities' perceptions of racism online and how this shapes attitudes towards individuals who confront. The aim of the present research is to examine how confrontation of racial discrimination through online interactions might impact perceptions of online prejudice and attitudes toward those who confront.

Being the Target of Discrimination

Increasing attention is being given to the impact that online racial discrimination can have on People of Color (POC; Bliuc et al., 2018). Similar to discrimination encountered offline, online racial discrimination occurs at comparable rates and is associated with increased feelings of depression and anxiety (Tynes, 2007; Tynes, Giang et al., 2008). When viewing racially offensive images posted on ostensible social media sites, Black individuals were more bothered as compared to their White counterparts. Further, Black individuals were more likely to take action against such behavior by "defriending" the original poster of the content, and/or posting responses that used language reflecting their negative perceptions of the post (Tynes & Markoe, 2010). While other research has shown both White individuals and POC to be equally offended by racist internet memes, POC who reported experiencing greater discrimination offline rated racist memes more negatively (Williams et al., 2016). Accordingly, POC may have unique reactions to racially discriminatory content online where their group membership and past experiences play an important role in their perceptions. Given the direct consequences of being targets of discrimination online, it is important to understand what shapes POC's perceptions and reactions to this type of content. Thus, we also aimed to examine whether past experiences with discrimination would influence perceptions of discriminatory content online.

Confronting Racial Discrimination

Choosing to confront, as opposed to remaining silent, is the volitional process during which one expresses disapproval of a person or group of people responsible for discriminating against others (Kaiser & Miller, 2004). Confrontation has been recognized as a "universally available prejudice reduction tool that does not rely on systematic intervention and that emphasizes the power of the individual" (Czopp & Ashburn-Nardo, 2012, pp. 176-177).

Participants confronted about their prejudiced attitudes reported lower prejudice in a subsequent task (Czopp et al., 2006). Similar findings occur in online contexts; Twitter users who committed racial harassment and were then sanctioned by a high-status user tended to reduce their use of racist language in future posts (Munger, 2017).

Type of Confrontation

Confrontations are often perceived as interactions that are imbued with hostility or anger, but in reality, confrontation can also take a passive form. For example, when discrimination takes an ambiguous form and/or is difficult to interpret, confrontations often take a passive form, such as distraction, to diffuse the situation (Reid & Dundes, 2017). However, passive confrontations may lead to less behavioral change of the perpetrator since their transgressions are not challenged, and thus may continue to perpetuate this behavior.

In contrast, direct aggressive confrontations are risky to enact. Aggressive confrontation may be perceived as violating social norms of politeness (Czopp et al., 2006) and threatens perpetrators’ self-image, which may instigate backlash (Baumeister et al., 1996). However, aggressive confrontation is often more effective in prejudice reduction, as transgressions are brought to perpetrators’ attention. While research has examined type of confrontation from the perpetrator’s perspective (Czopp et al., 2006), we were interested in how type of confrontation might moderate perceptions of racism online from the targets’ perspective. While Czopp et al.’s (2006) found that both hostile and calm confrontations were equally effective in reducing future instances of prejudice, we expect that this may not be the case when it comes to shaping perceptions of racist events. Particularly, posts and statements online can be highly ambiguous in nature (i.e., is this serious or a joke?) and therefore the reactions of others can serve as an extremely valuable cue to what is acceptable. Participants’ response to content online may be largely driven by descriptive norms (Cialdini et al., 1990) that are shaped by the dynamic nature of content online (i.e., responses to posts or threads). Thomas et al., (2019) found that when others reacted angrily to online clips of disparaging humor, participants then rated the content as less enjoyable. This suggests that others' reactions may shape individuals' perceptions of online content. Thus, we expect that when targets of discrimination see confronters reacting aggressively to a racist post, they should rate it as more offensive as compared to when confronters react more passively.

Group Membership of the Confronter

Despite confrontation being a promising avenue in which to combat racism, research has found that people rarely confront when witnessing instances of racism or prejudice (Dickter & Newton, 2013; Kawakami et al., 2009). One reason why people might be hesitant to confront discrimination is concerns about how they will be perceived by others. On the one hand, those who confront may be viewed more positively than those who do not (Kaiser et al., 2009), but it is also possible that they may be viewed more negatively (Zou & Dickter, 2013). Previous research shows mixed results on attitudes towards confronters. Kaiser et al. (2009) found that targets who confronted racism were evaluated more positively, while other research shows a backlash effect (Schultz & Maddox, 2013) against target members who chose to confront racism.

These inconsistent findings may largely be due to the group membership of the confronter. Confronters who were not the target of discrimination were perceived as more persuasive and favorable as compared to confronters belonging to the targeted group (Czopp & Monteith, 2003; Gulker et al. ,2013; Kutlaca et al., 2019; Rasinski & Czopp, 2010). Non-targets are often perceived as more persuasive because the action of confronting is not seen as advancing one's own interests (Eagly et al., 1978; Walster et al., 1966). Relatedly, much of this research suggests that high-status or majority groups have the power to define group norms (Aral & Walker, 2012; Munger, 2017; Paluck et al., 2016). *Group norm theory* posits that social norms should shape an individuals' perception and judgment (Sherif & Sherif, 1953). Thus, if majority group members define the norm in a context, we might expect that when POC are targets of racial discrimination, White individuals have the power to shift perceptions of racist content (i.e., if a White individual calls out racism, the transgression may be perceived as highly offensive). A confrontation by an outgroup member should shape how offensive a transgression is perceived to be (i.e., seen as more offensive if aggressive, and less offensive if passive).

Conversely, it may also be possible that target's (i.e., ingroup members) reaction (e.g., confrontation) to racist content online set the expectations for how people should react to the transgression. Minority group members are often subject to *targeted social referencing.* When the topic of discrimination is made salient, minority group members' opinions are perceived as "expert" and weighed much more heavily than those of majority group members (Crosby, 2015; Crosby & Monin, 2013; Crosby et al., 2008). If this is the case, we would expect that ingroup confronters and their type of response (aggressive vs. passive) should have more power to shape how offensive a transgression appears. However, it is unclear whether minority group members would similarly pay more attention to fellow ingroup members who confront in a scenario where they are targets of discrimination, as this research has primarily been conducted from the majority groups' perspective. There is some support that ingroup members would attend to other ingroup members; targets of discrimination are more likely to support ingroup confronters when they believe it calls to attention injustices against the group as a form of collective action (Kahn et al., 2016). Additionally, POC who highly identified with the targeted group also favored ingroup members who confronted more than their weakly identified counterparts (Kaiser et al., 2009). Thus, we were interested in examining how confrontation via the target/ingroup vs. non-target/outgroup, as well as strength of identity would shape perceptions of racist content online.

The Current Research

POC increasingly experience online racial discrimination and the majority of research examining confrontation has focused on perceptions of confrontation and confronters via the majority group perspective. Here we seek to extend the literature by examining how confrontation shapes POC's interpretation of online racism. Thus, the current studies examine whether type of confrontation (aggressive vs. passive) and group membership (ingroup vs. outgroup) of a confronter impacts how offensive a racist post is interpreted by minority group members. All measures (https://osf.io/ceqxd), manipulations, and exclusions are reported here and in their pre-registrations.

Study 1

In this study we examined how confrontation to online racial prejudice may alter Asian Americans' perceptions of offensiveness of a racist post, and whether the type of confrontation (aggressive vs. passive) or group membership of confronters (ingroup vs. outgroup) impacts those perceptions. We pre-registered this study (http://aspredicted.org/blind.php?x=3s4is3). In line with past research demonstrating that overt confrontations can be effective in reducing prejudice (Czopp et al., 2006), we expected that aggressive confrontations would be perceived as more "confrontational" as compared to passive confrontations; thus, we hypothesized that racist statements would be perceived as more offensive when accompanied by aggressive (vs. passive) confrontations.. In accordance with group norm theory, where behavior of others signals social norms for that context and shape observers' individuals' perceptions (Sherif & Sherif, 1953), we hypothesized that outgroup (i.e., White) confronters (vs. ingroup) would set social norms that racism unacceptable and signal that the racist post was more offensive.

Finally, we examined whether individual differences in racial identification and experiences with discrimination would moderate perceived offensiveness. In line with past work (e.g., Kaiser et al., 2009) we hypothesized that individuals who were more highly identified with their racial group would rate the posts as more offensive, and be more influenced by ingroup confronters' responses. Based on research demonstrating that POC who reported more experiences of discrimination offline rated racist internet memes are more offensive (Williams et al., 2016), we hypothesized that participants' experience with discrimination would similarly moderate perceptions of offensiveness.

Methods

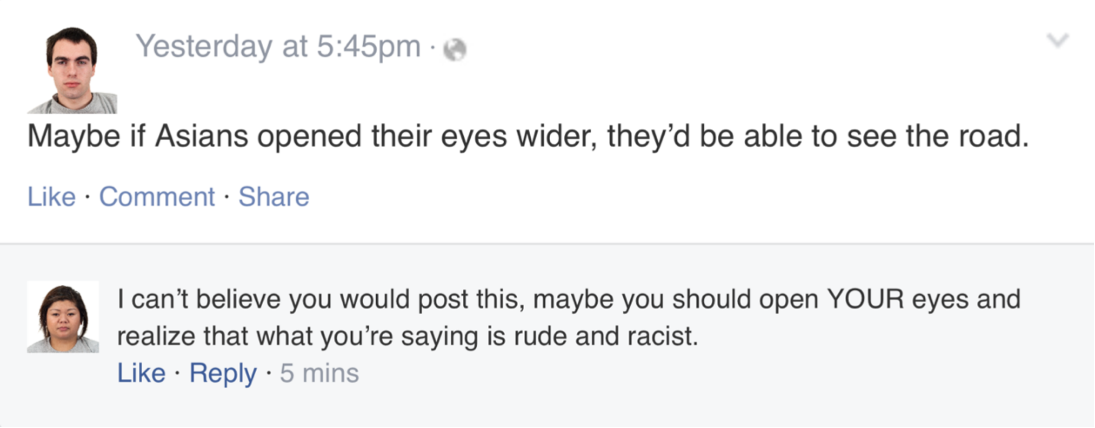
Seventy participants were recruited from the University of Hawai'i's human subjects participant pool, and received extra course credit as compensation for their time and effort. We aimed to collect at least 60 participants, based on an a priori power analyses for a repeated measures ANOVA with 4 measurements, with anticipated effect size of $f = .20$ and obtained power of 80%. Per our exclusion criteria, we removed the data of 9 participants who did not self-identify as monoracial East and Southeast Asian and/or took more +/- 2 standard deviations amount of time to complete the study from the analyses. Our final sample included 34 East Asian and 27 Southeast Asian participants. Of our 61 participants, 44 were female and 17 were male. Participants ranged from 17 to 67 years old ($M_{age} = 22.49$ years, $SD = 7.45$). A sensitivity power analyses assuming 80% power and alpha criteria of .05 was conducted to detect an effect size of .13 with sphericity assumption met at 1.00 and observed mean correlations among repeated measures at .62.

Procedure

Participants were asked to complete an online survey measuring perceptions of social interactions online. Participants were shown 16 mock Facebook posts, each of which included a racist statement accompanied by a confronting comment (see Figure 1). Facebook posts accompanied with comments were presented once at a time (randomized), and participants were asked to rate how offensive each original racist statement was. Following this, participants completed measures of experience with discrimination and ethnic identity (randomized).

Figure 1. Examples of stimuli. Panel A depicts an aggressive ingroup confronter. Panel B depicts a passive outgroup confronter.

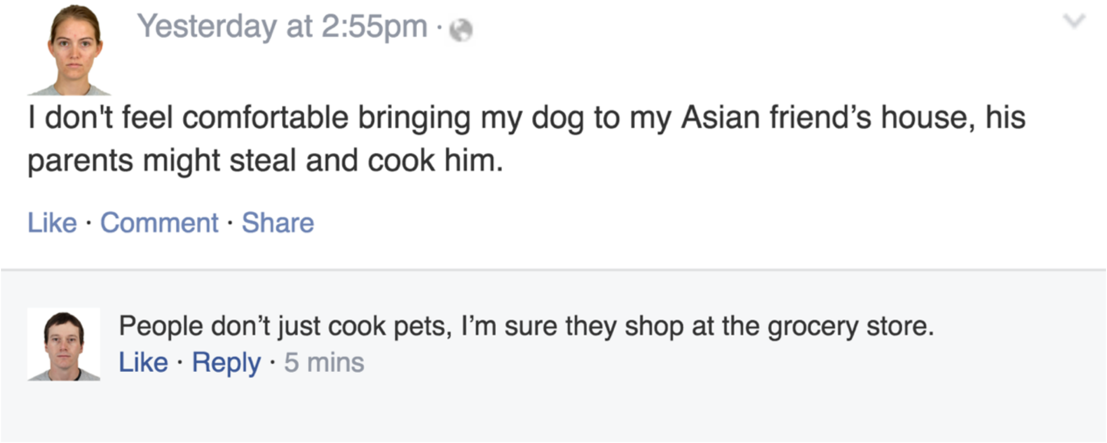
A



Maybe if Asians opened their eyes wider, they'd be able to see the road.

I can't believe you would post this, maybe you should open YOUR eyes and realize that what you're saying is rude and racist.

B


Yesterday at 2:55pm ·

I don't feel comfortable bringing my dog to my Asian friend's house, his parents might steal and cook him.

People don't just cook pets, I'm sure they shop at the grocery store.

Stimuli

Each status contained a racist statement made by a White outgroup member paired with a passive or aggressive confrontation, made by either an outgroup (White) or ingroup (Asian) confronter. Racist statements were pre-tested and matched on ratings to be highly offensive. Confrontations were pre-tested for how confrontational they were, such that aggressive confrontations were rated as high on confrontation whereas passive confrontations were rated low (see Supplemental Materials for additional information).

We obtained target faces from the Chicago Face Database (Ma et al., 2015). Our study included 24 White faces (16 for posts and 8 for confrontations) and 8 Asian faces that were matched in attractiveness and high in prototypicality for their respective race. Faces were also gender balanced, such that we had an equal amount of male and female faces. These stimuli were then input into a Facebook-like template that mimic the appearance of posts and comments seen on Facebook (see Figure 1) using Photoshop. In total, our stimuli included 4 posts that were accompanied with passive confrontations by Asian confronters, 4 posts that were accompanied with aggressive confrontations by Asian confronters, 4 posts that were accompanied with passive confrontations by White confronters, and 4 posts that were accompanied with aggressive confrontations by White confronters. All materials used across the studies are available here: https://osf.io/k4yg2/

Measures.

Offensiveness. Following each presentation of a post and comment, participants rated perceived offensiveness for each post on a 7-point Likert scale, from *1 = not offensive at all* to *7 = extremely offensive*.

Racial identity. To evaluate how strongly each participant identified with their ethnic group, we used the Identity Centrality Subscale (Luhtanen & Crocker, 1992). Participants responded to items such as: "The ethnic group I belong to is an important reflection of who I am." The 4-item subscale was measured on a 7-point Likert scale, from *1 = strongly disagree* to 7 = *strongly agre*e, with higher scores indicating greater strength in identity ($\alpha = .78$).

Past experiences with discrimination. To measure participants' past experience with discrimination, we used the Everyday Discrimination scale (Williams et al., 1997). Participants indicated how frequently they experienced discrimination in eight items, such as: "You are treated with less courtesy than other people" on a *1=never* to *6=almost everyday* scale. We omitted one item from the scale regarding intelligence, given that Asians are frequently associated with intelligence. Higher scores indicated greater past experience with discrimination ($\alpha = .83$).

Results

Analysis Strategy

We used the *lme4* (Bates et al., 2015) package on R to run a multilevel model for our repeated measures data. In line with current practice in this area of research, our pre-registration specifies a repeated measures ANOVA analyses. However, based on recommendations made by blind reviewers, we deviate[2] from this plan to report results from multilevel analyses that control multiple comparisons and within participant variability (see Gelman et al., 2012). We assumed random coefficients were correlated and used an unstructured covariance structure. All predictors were grand-mean centered prior to being included in the model. We included random effects for participants on Confrontation (Passive vs. Aggressive) and Group (Ingroup vs. Outgroup) as well as a random effect for Stimuli (see Judd et al., 2012). Confrontation and Group were both dummy coded, respectively, (0 = passive/outgroup, 1 = aggressive/ingroup). We initially ran models where each of our moderators (Identity and Discrimination) were entered in separately as fixed effects. There were no differences in results and so we report on the model with all predictors entered simultaneously in the model. See Table 1 for parameter estimates.

Perceptions of Offensiveness

There was a main effect of Confrontation, *p* < .001, where aggressively confronted statements were viewed as more offensive than passively confronted statements. However, this was qualified by an interaction between Confrontation and Group, *p* = .01. Simple effects analyses showed that when confronters were outgroup members, there were no differences in perceived offensiveness whether they were aggressive or passive, *b* = .13, *SE* = .08, 95% CI [-.04, .29], *t*(298) = 1.53, *p* = .13. However, when confronters were ingroup members, there was a significant difference in perceived offensiveness, *b* = .42, *SE* = .08, 95% CI [.26, .59], *t*(298) = 5.08, *p* < .001. Participants rated the post as significantly more offensive when it was accompanied by aggressive ingroup confronters (*M* = 5.99, *SD* = 1.35) as compared to when they were accompanied by passive ingroup confronters (*M* = 5.57, *SD* = 1.50).

Strength of Identity

[2] Analyses in line with our pre-registration are available in supplemental documents.

Participants' strength in their Asian identity did not relate to how offensive they perceived the post. Additionally, all interactions were non-significant.

Experiences with Discrimination

Similarly, participants' experiences with discrimination was not related to how offensive they perceived the post, nor did any of its interaction terms.

Table 1.

Parameter estimates for Study 1

Effect	Variance	SD	Estimate	*SE*	95% CI	*T*	*df*	*p*
Random effects								
Participants								
Intercept	1.78	1.33						
Confrontation	.01	.10						
Group	.005	.07						
Stimuli								
Intercept	.004	.06						
Residual	.82	.91						
Fixed effects								
Intercept			5.78	.18	5.44, 6.13	32.90	57.80	<.001
Confrontation (C)			.27	.06	.06, .16	4.61	89.60	<.001
Group (G)			-.02	.06	-.13, .10	-.28	354.80	.78
Identity (I)			.14	.13	-.12, .40	1.04	58	.30
Discrimination (D)			.03	.04	-.05, .10	0.68	58	.50
C X G			.30	.12	.07, .52	2.55	849	.01
C X I			-.04	.05	-.13, .05	-.77	89.6	.44
G X I			.01	.05	-.08, .10	.24	354.8	.81
C X D			.01	.01	-.02, .04	.69	89.6	.49
G X D			-.01	.01	-.03, .02	-.64	354.8	.52
C X G X I			.00	.09	-.17, .18	.02	845.00	.99
C X G X D			-.02	.03	-.07, .04	-.59	845.00	.56

Discussion

These results partially support our hypotheses, providing evidence that the type of confrontation matters for perceptions of online racism. Posts that were aggressively confronted were viewed as more offensive by our Asian American participants as compared to posts that were passively confronted. However, these results do not fully conform to our initial predictions regarding group membership of the confronter. Counter to our expectations, *group norm theory* (Sherif & Sherif, 1953) did not explain our results as there is no evidence that aggressive outgroup confronters influenced participants' perception of the initial racist post. Instead, participants viewed the racist post as most offensive when exposed to aggressive (vs. passive) ingroup confronters. These results suggest that Asian participants were referencing their fellow ingroup members to decide what was deemed offensive vs. not. Because this is not what we originally hypothesized, we aimed to replicate this in our next study. Lastly, we did not find any support for our moderations. Both strength of identity and experiences with discrimination did not significantly influence perceptions of racist statements.

Study 2

Given the unexpected results of Study 1, we aimed to replicate these findings in a second study. To improve ecological validity, we slightly altered the dependent variables in Study 2 to more closely mirror what an individual would experience on Facebook. This was done by offering participants the opportunity to report and respond to the post in addition to rating the perceived offensiveness of the post. Since experience with discrimination was not a significant predictor in ratings of offensiveness, we chose to omit this measure in all future studies[3]. We pre-registered this study (http://aspredicted.org/blind.php?x=jk575h) with hypotheses based on our findings from Study 1. We anticipated an interaction between confrontation type and group

[3] This decision was made based on the original pre-registered analyses (reported in the supplementary materials).

membership on offensiveness ratings, following a similar pattern in Study 1, such that ratings of offensiveness will significantly differ by confrontation type when confronters are ingroup members (i.e., when ingroup confronters are aggressive, ratings of offensiveness will be higher as compared to when ingroup confronters are passive). We also hypothesized a similar interaction for our new dependent variable, likelihood to report the statement to Facebook. For exploratory purposes, we also examined the likelihood of participants to respond to the post, as well as the content of the response (see Supplemental Materials for results on this measure).

Methods

Sixty-two participants were recruited from the University of Hawai'i's human subjects participant pool, and received extra course credit as compensation for their time and effort. We aimed to collect at least 60 participants, based on an a priori power analyses for a repeated measures ANOVA with 4 measures, with anticipated effect size of .20 and obtained power of 80%. Per our exclusion criteria, we dropped 2 participants who did not self-identify as monoracial East and Southeast Asian and/or took more +/- 2 standard deviations amount of time to complete the study. Our final sample included 35 East Asian and 25 Southeast Asian participants. Of our 60 participants, 38 were female and 22 were male. Participants ranged from 18 to 64 years old (M_{age} = 19.80 years, *SD* = 6.52). A sensitivity power analyses assuming 80% power and alpha criteria of .05 was conducted to detect an effect size of .15 with sphericity assumption met at 1.00 and observed mean correlations among repeated measures at .52.

Procedure

Participants followed the same procedure as Study 1, where they were shown 16 mock Facebook posts, each accompanied by a confronting comment (see Figure 1) identical to those in Study 1. Participants were presented a single Facebook post and comment, asked whether they would report the post to Facebook, were given the opportunity to respond to the post, then asked to rate the post's offensiveness, before being presented with the next stimuli. Lastly, participants completed a questionnaire on strength of racial identity.

Measures

Offensiveness ratings. See Study 1.

Likelihood to report. Mimicking Facebook's options to report content, participants were asked "Would you report the status to Facebook?" and were able to respond either "Yes" or "No". Responses were coded as 1 = Yes or 0 = No.

Racial identity subscale. We used the same identity centrality subscale as reported in Study 1 ($\alpha = .79$).

Results

Perceptions of Offensiveness

Using the same analytic approach as in Study 1, we found that both Confrontation and Group membership shaped perceived offensiveness of the racist post, *p*s < .003. Participants perceived the racist post as more offensive when accompanied with aggressive confrontations (*M* = 5.45, *SD* = 1.18) as opposed to passive confrontations (*M* = 5.12, *SD* = 1.33). Contrary to what we found in Study 1, this time participants perceived the racist post as more offensive when accompanied by outgroup confronters (*M* = 5.37, *SD* = 1.33) as compared to ingroup confronters (*M* = 5.21, *SD* = 1.31). Unlike Study 1, we found no interaction between Confrontation and Group, *p* = .50. Strength of identity was not related to perceived offensiveness, *p* = .09. See Table 2 for parameter estimates.

Likelihood to Report

Participants were more likely to report the racist post when accompanied by aggressive confrontations (*M* = 59%, *SD* = 40%) than when with a passive confrontation (*M* = 45%, *SD* = 40%). In other words, participants were approximately 3.29 times more likely to report a post when seeing an aggressive (vs. passive) confrontation. Participants' likelihood to report did not differ by confronter group membership, *p* = .17. Additionally, there was no interaction, *p* = .49. Strength of identity was not related to participants' likelihood to report the post, *p* = .15. See Table 3 for parameter estimates.

Table 2.

Parameter estimates for perceived offensiveness in Study 2

Effect	Variance	SD	Estimate	*SE*	95% CI	*t*	*df*	*p*
Random effects								
Participants								
Intercept	144.00	1.2						
Confrontation	.10	.31						
Group	.04	.20						
Stimuli								
Intercept	.00	.00						
Residual	.91	.96						
Fixed effects								
Intercept			5.22	.17	4.89, 5.55	31.34	62.50	<.001
Confrontation (C)			.29	.10	.10, .48	3.04	173.40	.003
Group (G)			-.20	.09	-.38, -.02	-2.19	264.80	.03
Identity (I)			.27	.16	-.04, .58	1.72	62.50	.09
C X G			.08	.12	-.16, .33	0.68	838	.50
C X I			-.008	.09	-.19, .17	-.08	173.40	.93
G X I			.04	.09	-.12, .21	.52	260.90	.60
C X G X I			.10	.12	-.13, .33	.86	836.00	.39

Table 3.

Parameter estimates for likelihood to report in Study 2

Effect	Variance	SD	Estimate	*SE*	95% CI	*exp(B)*	*z*	*p*
Random effects								
Participants								

Intercept	11.99	3.46						
Confrontation	.04	.20						
Group	.03	.17						
Stimuli								
Intercept	.12	.35						
Residual	1.00	1.00						
Fixed effects								
Intercept			-.33	.53	-1.39, .70	.72	-.63	.53
Confrontation (C)			1.19	.29	.63, 1.76	3.29	4.13	<.001
Group (G)			-.39	.28	-.94, .16	.68	-1.38	.17
Identity (I)			.67	.47	-.24, 1.59	1.96	1.44	.15
C X G			-.28	.41	-1.08, .52	.76	-.69	.49
C X I			.06	.26	-.44, .57	1.06	0.24	.81
G X I			.12	.25	-.37, .61	1.13	0.48	.63
C X G X I			-.06	.35	-.74, .63	.95	-0.16	.87

Discussion

The aim of Study 2 was to replicate our results on perceived offensiveness as well as extend our dependent variables to a behavioral response mirroring what would be encountered on Facebook – reporting and responding to posts. Again, we found partial support for our hypotheses. Replicating Study 1, confrontation type influenced perceived offensiveness; when posts were aggressively confronted, they were seen as more offensive and were more likely to be reported as compared to posts with passive confrontations. However, contrary to Study 1, we observed a reversal in the pattern of results for group membership. Supporting *group norm theory*, majority outgroup members' responses to racist posts shaped participants' perception of the post. In other words, our Asian American participants viewed posts confronted by outgroup members as more offensive, compared to posts confronted by ingroup members. This finding did not extend to our new dependent variable of likelihood to report the post, which was only modulated by type of confrontation.

One possible explanation for the observed pattern of results is that the procedure for this study created a more tangible context for interpreting and responding to posts. According to Construal Level Theory (CLT; Trope & Liberman, 2010), people's reactions differ depending on whether the context is hypothetical or actual. It is possible that in Study 1, evaluations of the post were on a hypothetical level, given that the task simply asked participants to view and rate posts. In Study 2, participants were given the opportunity to report and/or respond, which places them context where they are actively involved in the interaction. Thus, the threshold for determining whether a post is offensive enough to report may differ from than purely offering personal judgements, and such, people may look to majority outgroup members for cues on how to respond. These findings provide some evidence that perceptions of discriminatory statements online may be related to behaviors online (e.g., reporting and responding). However, it is unclear how the group membership of confronters influences these factors.

Study 3

Thus far, our findings have indicated that there is a robust main effect for confrontation type. When a racist post is confronted aggressively, it is seen as more offensive (Studies 1 and 2) and is more likely to be reported as offensive (Study 2). What remains unclear is the role of group membership in shaping targets' perception of racist statements. In Study 1, where participants were asked only to rate how offensive the post appears, we find that ingroup confronters shaped perceptions. However, in Study 2 when participants were asked to think less abstractly and decide whether to report the post or not before rating offensiveness, it was outgroup confronters who seemed to drive participants' perceptions. In an attempt to clarify this pattern of results, we again manipulated both confrontation type and group membership of confronters to examine the effects on offensiveness ratings of the original statement, replicating

Study 1 and 2. To extend this research, we examine the potential social cost of confronting by measuring participants' attitudes towards the confronter with an attitudes scale and warmth thermometer. We again provided participants with the opportunity to respond to the racist comments, similar to Study 2 (see Supplemental Materials for results on this exploratory measure).

We pre-registered this study (http://aspredicted.org/blind.php?x=7zz7ev) with hypotheses based upon our findings based on initial analyses in both Study 1 and 2. Given Study 2's results, we hypothesized that racist statements would be perceived as more offensive when confronted aggressively (vs. passively) and by an outgroup (vs. ingroup) member. But we anticipated this would come with little social cost. In line with previous research, we anticipated main effects would emerge such that aggressive (vs. passive) and ingroup (vs. outgroup) confronters would be viewed more favorably. However, we expected that these two would interact, such that outgroup members would be perceived just as favorably as ingroup members, but only when confronting in an aggressive manner.

Methods

Seventy-four participants were recruited from the University of Hawai'i's human subjects participant pool, and received extra course credit as compensation for their time and effort. We aimed to collect at least 60 participants, based on an a priori power analyses for a repeated measures ANOVA with 4 measures, with anticipated effect size of .20 and obtained power of 80%. Per our exclusion criteria, we dropped 13 participants who did not self-identify as monoracial East and/or Southeast Asian and/or took more +/- 2 standard deviations amount of time to complete the study. Our final sample included 30 East Asian and 27 Southeast Asian, and 4 biracial Southeast Asian/East Asian participants. Of our 61 participants, 33 were female and 28 were male. Participants ranged from 18 to 44 years old (M_{age} = 20.5 years, SD = 4.78). A sensitivity power analyses assuming 80% power and alpha criteria of .05 was conducted to detect an effect size of .12 with sphericity assumption met at 1.00 and observed mean correlations among repeated measures at .69.

Procedure

Participants followed a similar procedure to Study 2, where they were shown 16 mock Facebook posts, each of which was accompanied by a confronting comment (see Figure 1). Participants were given an opportunity to respond to the post, rate how offensive the post appeared, and asked about their attitudes and feelings towards the confronter[4]. Lastly, participants were asked to complete a questionnaire on strength of racial identity.

Measures

Offensiveness ratings. See Study 1.

Attitudes towards the confronter. To evaluate participants' attitudes towards each confronter, we used 4 items from Kaiser et al. (2009), and asked participants to indicate how they felt about the confronter. Participants responded to items, such as ''I would want the individual as a very close friend", on a 7-point scale, from *0 = strongly disagree to 6 = strongly agree* (α = .98).

Racial identity subscale. We used the same identity centrality subscale as reported in Study 1 (α = .85).

Results

Perceptions of Offensiveness

[4] Our measures of attitudes towards the confronter and feeling thermometers were highly correlated, r = .90, p < .001, and demonstrate similar results. Analyses for feeling thermometers are reported in the supplemental materials.

We again found that both Confrontation and Group membership shaped perceived offensiveness of the racist post, *p*s < .03. Participants perceived the racist post as more offensive when accompanied with aggressive confrontations (*M* = 5.88, *SD* = 1.34) as opposed to passive confrontations (*M* = 5.58, *SD* = 1.36). Replicating Study 2, participants perceived the racist post as more offensive when accompanied by outgroup confronters (*M* = 5.78, *SD* = 1.37) as compared to ingroup confronters (*M* = 5.68, *SD* = 1.33). Again, we found no interaction between Confrontation and Group, *p* = .12. However, this time strength of identity was related to perceived offensiveness, *p* = .004, such that those who held stronger Asian identities perceived the post as more offensive. See Table 4 for parameter estimates.

Attitudes Towards the Confronter

As hypothesized, there was a main effect for Confrontation, *p* < .001, such that attitudes towards the confronter was more positive when the confronter was aggressive (vs. passive). As expected, we also found a main effect for Group on attitudes towards the confronter, *p* < .001, such that participants felt more positive towards ingroup confronters as compared to outgroup confronters. Lastly, as anticipated, we found a Confrontation X Group interaction for attitudes towards the confronter, *p* = .001. Simple effects analyses found that when confrontations were aggressive, both ingroup and outgroup confronters were perceived positively, *b* = -.06, *SE* = .09, 95% CI [-.25, .13], *t*(202) = -.64, *p* = .53. However, when confrontations were passive, participants viewed passive ingroup members more favorably as compared to passive outgroup members, *b* = .35, *SE* = .09, 95% CI [.17, .54], *t*(202) = 3.74, *p* <.001. We also found an interaction between Group X Identity, *p* = .02. Simple effects analyses show that for participants who were weak (-1 *SD*) in Asian identification did not differ in their attitudes towards ingroup vs. outgroup confronters, *b* = .0007, *SE* = .10, 95% CI [-.19, .19], *t*(60) = .008, *p* = .99. However, participants who had strong Asian identities (+1 *SD*) viewed ingroup confronters more favorably than outgroup confronters, *b* = .29, *SE* = .10, 95% CI [.10, .48], *t*(60) = 3.01, *p* = .004. See Table 5 for parameter estimates.

Table 4.

Parameter estimates for perceived offensiveness in Study 3

Effect	Variance	SD	Estimate	*SE*	95% CI	*t*	*df*	*p*
Random effects								
Participants								
Intercept	1.55	1.24						
Confrontation	.02	.14						
Group	.02	.14						
Stimuli								
Intercept	.003	.06						
Residual	.82	.91						
Fixed effects								
Intercept			5.66	.17	5.33, 5.99	33.18	63.10	<.001
Confrontation (C)			.22	.08	.06, .39	2.66	356.80	.008
Group (G)			-.18	.08	-.34, -.02	-2.18	413.20	.03
Identity (I)			.39	.13	.14, .65	3.01	64.00	.004
C X G			.18	.12	-.05, .40	1.54	861.00	.12
C X I			-.05	.06	-.18, .07	-.82	356.80	.41
G X I			.05	.06	-.07, .18	.82	413.20	.41
C X G X I			.0007	.09	-.17, .18	.008	861.00	.99

Table 5.

Parameter estimates for attitudes towards confronter in Study 3

Effect	Variance	SD	Estimate	*SE*	95% CI	*t*	*df*	*p*
Random effects								
Participants								
Intercept	1.60	1.27						
Confrontation	1.06	1.03						
Group	.03	.18						
Stimuli								
Intercept	.006	.08						

Residual	1.02	1.01						
Fixed effects								
Intercept			3.40	.18	3.06, 3.75	19.19	61.60	<.001
Confrontation (C)			.91	.16	.60, 1.22	5.7	85.40	<.001
Group (G)			.35	.09	.17, .53	3.74	201.60	<.001
Identity (I)			-.17	.13	-.43, .09	-1.25	64.30	.21
C X G			-.41	.13	-.66, -.16	-3.19	801.00	.001
C X I			.19	.12	-.05, .43	1.53	85.40	.13
G X I			.17	.07	.02, .31	2.30	201.60	.02
C X G X I			-.11	.10	-.30, .09	-1.10	801.00	.27

Discussion

We replicated our findings from Study 2, such that racist posts accompanied with aggressive confrontations and outgroup confronters were rated as more offensive as compared to statements with passive confrontations and ingroup confronters. Additionally, we found a main effect for strength of identity, such that participants who had a strong Asian identity found the posts overall more offensive.

When examining how these factors impacted attitudes towards the confronter, we found that participants felt more positively towards confronters who were aggressive (vs. passive) and ingroup (vs. outgroup) members. A confrontation type x group membership interaction also emerged. When confronters were aggressive, they were viewed positively regardless of their racial group membership. However, when confronters were passive, outgroup confronters were viewed more negatively than ingroup confronters. Despite outgroup confronters' power to shape perceptions of the racist post, they were only seen positively if they confronted in an aggressive manner.

Somewhat surprising given the pattern of results from the previous studies, in Study 3 participants' own identity levels impacted their evaluations of racist posts. Here, participants who identified strongly with their group viewed posts as more offensive, regardless of input related to confrontation. This maps onto previous work demonstrating that highly identified individuals react negatively towards events that reflect poorly on their group and (Abrams et al., 2000; Branscombe et al., 1993). Seeing any racist post towards their group would be threatening to a highly identified individual, and such, these individuals would find the post more offensive regardless of how others responded. As anticipated, we also observed an influence of racial identity on evaluations of confronters. Highly identified Asian participants favored ingroup confronters more than outgroup confronters, regardless of the manner in which they confronted the transgression. However, unlike other work in this area (e.g., Vaccarino & Kawakami, in press), backlash against the target for passive behavior was not found. Rather, highly identified participants only evaluated the outgroup confronter negatively for responding passively. These results are in line with previous research investigating the moderating role racial identification on attitudes towards confronters of discrimination (Kaiser et al., 2009) where highly identified targets favored ingroup confronters. The current findings illustrate the importance of identity when examining evaluations of confronters.

General Discussion

Across three studies we demonstrate that aggressive confrontations shift perceptions of racists statements online. When racial discrimination was aggressively confronted, participants were more likely to perceive the perpetrating statement as significantly more offensive. Contrary to our expectations, our findings were mixed for group membership. While we initially found that ingroup confronters had more power to shape perceptions of racists posts (Study 1), we found the opposite was true for Studies 2 and 3. One possible explanation for these discrepancies is the context in which these studies were situated. According CTL (Trope & Liberman, 2010), Study 1 could be characterized as a more abstract context in which participants were asked to make judgments about posts. By contrast, Studies 2 and 3 asked participants to report and/or respond to Facebook posts before judging the offensiveness of the content. By asking participants to participate in the Facebook thread, the context in which they made evaluations about the post were less psychologically distant. Given the higher ecological validity of Studies 2 and 3, it is most likely that outgroups, and not ingroups set the social norms in relation to what is deemed offensive online.

Aggressive confrontation not only shifted perceptions, but increased the likelihood of the racist post being reported. These findings illustrate the importance of how POC perceptions of online content may be related to subsequent behavior that reduces the presence of online discrimination. Banning offensive content, such as racist and discriminatory remarks, is an important step in communicating that these behaviors are not acceptable, even in an online platform. Given that research has found that people experience similar psychological health consequences from both off- and on-line content (Tynes, 2007; Tynes et al., 2008), reducing the instances of discrimination online is imperative. Seeing others' outrage in response to a post may motivate one to report or ban the behavior and reduce the likelihood of others seeing this type of harmful content online.

Importantly, the benefits of aggressive confrontation do not come at social cost to the confronter. In Study 3, aggressive confronters were perceived more favorably as compared to passive confronters, regardless of group membership. It was only when outgroup members were passive in their confrontation that they received backlash, and were evaluated significantly more negatively than their passive ingroup counterparts. Since the outgroup members and perpetrator of the racist statement were both White individuals, it brings to question whether White people are expected to “police” or moderate their fellow ingroup members. Other research in this area has shown that targets are expected to speak out against discrimination towards their group, yet they suffer the greatest consequences as a result of doing so (Vaccarino & Kawakami, in press). Furthermore, there is evidence to suggest that majority group members are less likely to speak out against discrimination unless they receive reassurance (typically from the targets of discrimination) that an event is offensive or warrants response (Crosby, 2015). However, much of this research has been primarily conducted from the majority groups’ perspective, whereas in the current study we examine minority group perceivers, who are the targets of discrimination in this context. Here we find initial evidence that POC support White individuals who aggressively (but not passively) confront online racial discrimination.

Lastly, we found that strength of identity moderated evaluations of the confronter, such that strongly identified individuals demonstrated stronger ingroup favoritism as compared to weakly identified individuals, regardless of the confrontation type. It is possible that ingroup confronters, regardless of their confrontation are seen as fellow targets of discrimination, and therefore viewed positively for those who are highly identified with their group.

Overall, the present research has important implications for how confrontation shapes POC perceptions of online discrimination. Confronting discrimination has been touted as a clear way to combat prejudice (Czopp & Ashburn-Nardo, 2012), but there may be costs and benefits to the manner in which people choose to confront. Particularly, within social media online, we may be exposed to more instances of blatant prejudice. Yet these do not occur within a vacuum, the dynamic nature of social media includes dialogue from varied individuals and with distinct opinions. It is important to understand how perceptions of discrimination and prejudice are impacted by these dynamic factors. The present research suggests that when people react to racists content online by aggressively reprimanding them, it may signal to others that this behavior is not acceptable. Similar to findings by Thomas et al., (2019), evaluations of content viewed online are largely shaped by the reactions of others. Reactions, in the form of confrontation, may set norms for what is perceived as acceptable vs. unacceptable, which may be especially important online.

Limitations and Future Directions

Although the findings across three studies offer consistent evidence that confrontation type predicts how offensive racist statements are perceived to be, this effect may be limited to online conversations between White and Asian individuals, specifically where Whites are the perpetrators of discriminatory comments directed towards Asian individuals. Additional research should also be conducted to evaluate the role of perpetrator group membership in greater depth (for example, do the same effects exist if the perpetrators are other minority group members). Furthermore, given the mixed results regarding group membership across our studies, more work is needed to disentangle whether targets of discrimination are more likely to socially tune to their ingroup (i.e., targeted social referencing; Crosby, 2015) or adhere to majority group norms (i.e., group norm theory; Sherif & Sherif, 1953).

Past research has shown that more aggressive forms of confrontation (as opposed to passive confrontation) is more likely to lead to prejudice reduction (Czopp et al., 2006), and thus we chose to specifically look at aggressive vs. passive confrontation, as opposed to confrontation vs. no confrontation like past studies have done (Ashburn-Nardo et al., 2014; Good et al., 2012; Kaiser et al., 2009; Rattan & Dweck, 2010). While we focused on the type of confrontation, it remains unclear from our findings if passive confrontation is any worse than no confrontation at all. Future research should examine whether passive confrontation is any more beneficial than no confrontation. Moreover, there may be instances in which passive confrontation could actually be beneficial (e.g., a situation where the transgression is ambiguous or dangerous). Future research should examine how contextual factors might dictate the best approach to confronting prejudice. The type of confrontation has the potential to meaningfully impact the direction of a conversation, whether it derails the conversation away from prejudice, or if it helps to condone prejudiced attitudes. Understanding what type of confrontation leads to shifts in perceptions, attitudes, and behaviors may help us figure out what makes a good ally. Allies can confront discriminatory comments in a way that is aggressive, and reprimands discriminatory behavior, and help set the norms in conversations online.

Conclusion

This research extends our understanding of how confrontation may operate in online contexts, and its impact on perceptions and attitudes. Specifically, we find that aggressive confrontation, as opposed to passive, make original transgressions online appear more offensive. Furthermore, we find that aggressive confronters are perceived more favorably by targets of discrimination. Importantly, this highlights the importance of how confrontation is performed. While passive confrontations may still be considered a form of confrontation, if it distracts from directly condemning racist attitudes, may still signal acceptability. Confrontation can signal social norms on whether expression of prejudice is acceptable or not (Nelson et al., 2011). Together this research suggests that confronting discrimination in a more direct manner may reduce the expression of bias in the future, have positive outcomes for confronters and targets alike, and signal norms in society of what is acceptable behavior.

References

Abrams, D., Marques, J. M., Bown, N., & Henson, M. (2000). Pro-norm and anti-norm deviance within and between groups. *Journal of Personality and Social Psychology*, *78*(5), 906–912. https://doi.org/10.1037/0022-3514.78.5.906

Aral, S., & Walker, D. (2012). Identifying influential and susceptible members of social networks. *Science*, *337*(6092), 337-341. https://doi.org/10.1126/science.1215842

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A., & Goodwin, S. A. (2014). Do you say something when it's your boss? The role of perpetrator power in prejudice confrontation. *Journal of Social Issues, 70*(4), 615-636. https://doi.org/10.1111/josi.12082

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi/org/10.18637/jss.v067.i01

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review, 103*(1), 5. https://doi.org/10.1037/0033-295X.103.1.5

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

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-group or out- group extremity: Importance of the threatened social identity. *Personality and Social Psychology Bulletin*, *19*(4), 381–388. https://doi.org /10.1177/0146167293194003

Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015-1026. https://doi.org/10.1037/0022-3514.58.6.1015

Cleland, J. (2014). Racism, football fans, and online message boards: How social media has added a new dimension to racist discourse in English football. *Journal of Sport and Social Issues, 38*(5), 415-431. https://doi.org/10.1177/0193723513499922

Crosby, J. R. (2015). The silent majority: Understanding and increasing majority group responses to discrimination. *Social and Personality Psychology Compass*, *9/10*, 539–550. http://doi.org/10.1111/spc3.12196

Crosby, J. R., & Monin, B. (2013). How the opinions of racial minorities influence judgments of discrimination. *Basic and Applied Social Psychology*, *35*(4), 334–345. http://doi.org/10.1080/01973533.2013.803963

Crosby, J. R., Monin, B., & Richards, D. (2008). Where do we look during potentially offensive behavior? *Psychological Science*, *19*(3), 226–228. http://doi.org/10.1111/j.1467-9280.2008.02072.x

Czopp, A. M., & Ashburn-Nardo, L. (2012). Interpersonal confrontations of prejudice. In D. W. Russell & C. A. Russell (Eds.), *The psychology of prejudice: Interdisciplinary perspectives on contemporary issues.* Hauppauge, NY: Nova Science Publishers, Inc.

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin,* 29, 532-544. https://doi.org/10.1177/0146167202250923

Czopp, A. M., Monteith, M. J., & Mark, A. Y. (2006). Standing up for a change: Reducing bias through interpersonal confrontation. *Journal of Personality and Social Psychology, 90*(5), 784-803. https://doi.org/10.1037/0022-3514.90.5.784

Daniels, J. (2012). Race and racism in internet studies: A review and critique. *New Media and Society, 15*(5), 695-719. https://doi.org/10.1177/1461444812462849

Dickter, C. L., & Newton, V. A. (2013). To confront or not to confront: Non-targets' evaluations of and responses to racist comments. *Journal of Applied Social Psychology*, 43, E262-E275. https://doi.org/10.1111/jasp.12022

Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and social Psychology, 36*(4), 424. https://doi.org/10.1037/0022-3514.36.4.424

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. https://doi.org/10.1080/19345747.2011.618213

Good, J. J., Moss-Racusin, C. A., & Sanchez, D. T. (2012). When do we confront? Perceptions of costs and benefits predict confronting discrimination on behalf of the self and others. *Psychology of Women Quarterly, 36*(2), 210-226. https://doi.org/10.1177/0361684312440958

Gulker, J. E., Mark, A. Y., & Monteith, M. J. (2013). Confronting prejudice: The who, what, and why of confrontation effectiveness. *Social Influence, 8*(4), 280-293. https://doi.org/10.1080/15534510.2012.736879

Jakubowicz, A., Dunn, K., Mason, G., Paradies, Y., Bliuc, A. M., Bahfen, N., ... & Connelly, K. (2017). How Cyber Users Experience and Respond to Racism: Evidence from an Online

Survey. In *Cyber Racism and Community Resilience* (pp. 65-94). Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-64388-5_3

Johns, A., & McCosker, A. (2015). Social media conflict: Platforms for racial vilification, or acts of provocation and citizenship? *Communication, Politics & Culture, 47*(3), 44.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. http://doi.org/10.1037/a0028347

Kahn, K. B., Barreto, M., Kaiser, C. R., & Rego, M. S. (2016). When do high and low status group members support confrontation? The role of perceived pervasiveness of prejudice. British *Journal of Social Psychology, 55*(1), 27-43. https://doi.org/10.1111/bjso.12117

Kaiser, C. R., Hagiwara, N., Malahy, L. W., & Wilkins, C. L. (2009). Group identification moderates attitudes toward ingroup members who confront discrimination. *Journal of Experimental Social Psychology, 45*(4), 770-777. https://doi.org/10.1016/j.jesp.2009.04.027

Kaiser, C. R., & Miller, C. T. (2004). A stress and coping perspective on confronting sexism. *Psychology of Women Quarterly, 28*(2), 168-178. https://doi.org/10.1111/j.1471-6402.2004.00133.x

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Misreading affective and behavioral responses to racism. *Science*, *323*(5911), 276-278. https://doi.org/10.1126/science.1164951

Kutlaca, M., Becker, J., & Radke, H. (2019). A hero for the outgroup, a black sheep for the ingroup: Societal perceptions of those who confront discrimination. *Journal of Experimental Social Psychology*. *88*. https://doi.org/10.1016/j.jesp.2019.103832

Luhtanen, R., & Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin, 18*(3), 302-318. https://doi.org/10.1177/0146167292183006

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods, 47*(4), 1122-1135. https://doi.org/10.3758/s13428-014-0532-5

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

Nelson, J. K., Dunn, K. M., & Paradies, Y. (2011). Bystander anti-racism: A review of the literature. *Analyses of Social Issues and Public Policy, 11*(1), 263-284. https://doi.org/10.1111/j.1530-2415.2011.01274.x

Paluck, E. L., Shepherd, H., & Aronow, P. M. (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, *113*(3), 566-571. https://doi.org/10.1073/pnas.1514483113

Pew Research Center, (2016). "Social Media Conversations About Race" Available at: http://www.pewinternet.org/2016/08/15/social-media-conversations-about-race/

Rasinski, H. M., & Czopp, A. M. (2010). The effect of target status on witnesses' reactions to confrontations of bias. *Basic and Applied Social Psychology, 32*(1), 8-16. https://doi.org/10.1080/01973530903539754

Rattan, A., & Dweck, C. S. (2010). Who confronts prejudice? The role of implicit theories in the motivation to confront prejudice. *Psychological Science, 21*(7), 952-959. https://doi.org/10.1177/0956797610374740

Rauch, S. M., & Schanz, K. (2013). Advancing racism with Facebook: Frequency and purpose of Facebook use and the acceptance of prejudiced and egalitarian messages. *Computers in Human Behavior, 29*(3), 610-615. https://doi.org/10.1016/j.chb.2012.11.011

Reid, A., & Dundes, L. (2017). Bystander programs: accommodating or derailing sexism? *Behavioral Sciences, 7*(4), 65. https://doi.org/10.3390/bs7040065

Schultz, J. R., & Maddox, K. B. (2013). Shooting the messenger to spite the message? Exploring reactions to claims of racial bias. *Personality and Social Psychology Bulletin, 39*(3), 346-358. https://doi.org/10.1177/0146167212475223

Sherif, M., & Sherif, C. W. (1953). *Groups in harmony and tension; an integration of studies of intergroup relations.* Harper & Brothers.

Thomas, E. F., McGarty, C., Spears, R., Livingstone, A. G., Platow, M. J., Lala, G., & Mavor, K. (2019). “That's not funny!” Standing up against disparaging humor. *Journal of Experimental Social Psychology*, *86*, 103901. http://doi.org/10.1016/j.jesp.2019.103901

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. http://doi.org/10.1037/a0018963

Tynes, B. M. (2007). Internet safety gone wild? Sacrificing the educational and psychosocial benefits of online social environments. *Journal of Adolescent Research, 22*(6), 575-584. https://doi.org/10.1177/0743558407303979

Tynes, B. M., Giang, M. T., Williams, D. R., & Thompson, G. N. (2008). Online racial discrimination and psychological adjustment among adolescents. *Journal of Adolescent Health, 43*(6), 565-569. https://doi.org/10.1016/j.jadohealth.2008.08.021

Tynes, B. M., & Markoe, S. L. (2010). The role of color-blind racial attitudes in reactions to racial discrimination on social network sites. *Journal of Diversity in Higher Education, 3*(1), 1. https://doi.org/10.1037/a0018683

Vaccarino, E., & Kawakami, K. (in press) In the office or at the gym: The impact of confronting sexism in specific contexts on support for confrontation and perceptions of others. *Self & Identity.*

Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology, 2*(4), 325-342. https://doi.org/10.1016/0022-1031(66)90026-6

Williams, A., Oliver, C., Aumer, K., & Meyers, C. (2016). Racial microaggressions and perceptions of Internet memes. *Computers in Human Behavior*, 63, 424-432. https://doi.org/10.1016/j.chb.2016.05.067

Williams, D. R., Yu, Y., Jackson, J. S., & Anderson, N. B. (1997). Racial differences in physical and mental health: Socio-economic status, stress and discrimination. *Journal of Health Psychology, 2*(3), 335-351. https://doi.org/10.1177/135910539700200305

Winter, S., Brückner, C., & Krämer, N. C. (2015). They came, they liked, they commented: Social influence on Facebook news channels. *Cyberpsychology, Behavior, and Social Networking, 18*(8), 431-436. https://doi.org/10.1089/cyber.2015.0005

Zou, L. X., & Dickter, C. L. (2013). Perceptions of racial confrontation: The role of color blindness and comment ambiguity. *Cultural Diversity and Ethnic Minority Psychology, 19*(1), 92-96. https://doi.org/10.1037/a0031115

Supplementary Material

Stimuli Pre-testing

Racist Posts

Participants ($N = 20$) were individually presented with each post and asked to rate "How offensive do you find this statement?" on a scale of 1 *(not at all)* to 7 *(extremely)*. The selected statements (16 total) were selected because they were rated as being extremely offensive, with each participant responding with 5 or above ($M = 6.38$, $SD = .42$). Statements were then randomly assigned to confrontation type and group conditions. A 2 (Confrontation Type: aggressive vs. passive) X 2 (Group Membership: ingroup vs outgroup) ANOVA pre-testing scores revealed no significant differences in perceived offensiveness of racist posts, *p*s > .30. See Table 1 for means and standard deviations across all factors.

Confrontations

The same group of participants individually rated each comment with its related post on "How confrontational is this response?" from 1 (*not at all)* to 7 (*extremely*). Comments receiving a rating of 5 or above were considered to be highly confrontational and were categorized as aggressive confrontations ($M = 5.64$, $SD = .28$). Comments receiving a rating of 2 or below were deemed as low in confrontation were categorized as passive confrontations ($M = 1.38$, $SD = .32$). A 2 (Confrontation Type: aggressive vs. passive) X 2 (Group Membership: ingroup vs outgroup) ANOVA pre-testing scores revealed that there was a significant difference in how confrontational comments were across confrontation type, $F(1, 12) = 710.45$, $p < .001$. Importantly, comments did not differ in how confrontational they were across group membership, $F(1, 12) = .06$, $p = .82$. Similarly, there was no interaction, $F(1, 12) = .12$, $p = .74$. See Table 2 for means and standard deviations across all factors.

Table 1.

Mean ratings of offensiveness of posts by confrontation and group

Type of comment	Ingroup	Outgroup

Aggressive	6.39 (.56)	6.59 (.18)
Passive	6.35 (.44)	6.17 (.45)

Note. Standard deviations are in parentheses.

Table 2.

Mean ratings of offensiveness of comments by confrontation and group

Type of comment	Ingroup	Outgroup
Aggressive	5.65 (.25)	5.63 (.35)
Passive	1.33 (.40)	1.43 (.27)

Note. Standard deviations are in parentheses.

Study 1: Pre-registered analyses

In this study we examined how confrontation to online racial prejudice may alter perceptions of offensiveness of the racist post, and whether the type of confrontation (aggressive vs. passive) or group membership of confronters (ingroup vs. outgroup) impacts those perceptions. We pre-registered this study (http://aspredicted.org/blind.php?x=3s4is3) with the following hypotheses. H1: We anticipated racist statements would be perceived as more offensive when presented with an aggressive (vs. passive) confronting comment, H2: racist statements will be perceived as more offensive when presented with an outgroup (vs. ingroup) confronter, and H3: racist statements confronted with an aggressive confrontation (vs. passive) will be perceived as more offensive when the confronter is an outgroup (vs. ingroup) member. H4: Additionally, we hypothesized strength of racial identification will moderate this relationship, such that those with stronger racial identification will view racist statements aggressively confronted by an outgroup member as more offensive, and lastly, H5: experience with discrimination will moderate this relationship, such that those experiencing more racial discrimination offline will view racist statements aggressively confronted by an outgroup member as more offensive.

Results

Perceptions of Offensiveness

In order to examine the impact of confrontation type and group membership on how offensive online racist posts were perceived (H1-H3), we conducted a 2 (Confrontation: Passive vs. Aggressive) X 2 (Group: Ingroup vs. Outgroup) repeated measures ANOVA on ratings of offensiveness. There was a main effect for confrontation, $F(1, 60) = 24.27$, $p < .001$, $\eta_p{}^2 = .29$, where statements were reported as more offensive when accompanied with an aggressive ($M = 5.92$, $SD = 1.36$) confrontation as compared to a passive ($M = 5.65$, $SD = 1.38$) confrontation. There was no main effect for group, $F(1, 60) = .13$, $p = .72$. This interaction was not significant, $F(1, 60) = 4.18$, $p = .05$, $\eta_p{}^2 = .07$.

Strength of Identity

In order to test whether identity moderated the above effect (H4) we conducted a hierarchical linear regression on our dependent variable of offensiveness with the following predictors and their interaction terms: group (effect coded as 1 = outgroup vs. -1 = ingroup), confrontation (effect coded as 1 = aggressive vs. -1 = passive) and strength of identity (grand-mean-centered). In the first model of our regression, we included all predictors. In the second model of our regression, we included all two-way interactions. In the last model of our regression, we included the three-way interaction. Our first model accounted for 3.19% of variance, and the only predictor found to be significantly related to ratings of offensiveness was strength of identity, such that those who reported greater strength of identity, rated statements as more offensive, β= .15, p = .02. Our second and third model did not significantly account for any more variance, as compared to the first model, *ps* > .87, and none of our additional predictors significantly related to offensiveness ratings.

Experience with Discrimination

In order to test whether experience with discrimination offline moderated the above effect (H5) we conducted a hierarchical linear regression on our dependent variable of offensiveness with the following predictors and their interaction terms: group (effect coded as 1 = outgroup vs. -1 = ingroup), confrontation (effect coded as 1 = aggressive vs. -1 = passive) and experience with discrimination (grand-mean-centered). In the first model of our regression, we included all predictors. In the second model of our regression, we included all two-way interactions. In the last model of our regression, we included the three-way interaction. Our first model accounted for 2.25% of variance, and none of our predictors significantly related to ratings of offensiveness. Our second and third model did not significantly account for any more variance, as compared to the first model, *ps* > .85.

Discussion

Our pre-registered analyses and multilevel analyses differ in a few important ways. While our main effect for confrontation type is consistent, we fail to find a significant interaction in our pre-registered analyses. Furthermore, we do see a relationship emerge for strength of identity and perceptions of offensiveness in our pre-registered analyses. This finding led us to include our identity measurement in future studies. Replicating our multilevel analyses, we do not find any relationship between experience with discrimination and perceived offensiveness.

Study 2: Pre-registered analyses

We pre-registered this study (http://aspredicted.org/blind.php?x=jk575h) with the following hypotheses, based on our findings in Study 1. H1: Replicating Study 1, we predicted racist statements will be perceived as more offensive when presented with an aggressive confrontation. H2: We anticipated an interaction between group membership and confrontation type (aggressive vs. passive) on offensiveness ratings, following a similar pattern in Study 1, such that ratings of offensiveness will significantly differ by confrontation type when confronters are ingroup members (i.e., when ingroup confronters are aggressive, ratings of offensiveness will be higher as compared to when ingroup confronters are passive). H3: Additionally, replicating Study 1, strength of racial identification will be related to ratings of offensiveness, such that those with stronger racial identification will rate all statements as more offensive, and lastly, H4: we hypothesized a similar interaction as in H2, for our new dependent variable, likelihood to report the statement to Facebook.

Results

Perceptions of Offensiveness

To test H1 and H2, we conducted a 2 (Confrontation: Aggressive vs. Passive) X 2 (Group: Ingroup vs. Outgroup) repeated measures ANOVA on ratings of offensiveness. As hypothesized, we found a main effect for confrontation type, $F(1, 59) = 21.06$, $p < .001$, $\eta_p^2 = .26$, such that statements accompanied by aggressive confrontations ($M = 5.45$, $SD = 1.18$) were rated as more offensive than those with passive confrontations ($M = 5.12$, $SD = 1.33$). Unexpectedly, we found a main effect for group membership, $F(1, 59) = 5.63$, $p = .02$, $\eta_p^2 = .09$. Statements accompanied with outgroup confronters ($M = 5.37$, $SD = 1.19$) were rated as more offensive compared to those with ingroup confronters ($M = 5.21$, $SD = 1.31$). There was no significant interaction between confrontation type and group membership, $F(1, 59) = .62$, $p = .43$.

Likelihood to Report

Following H4, we expected similar effects for likelihood to report as we found for ratings of offensiveness. We conducted a 2 (Confrontation: Aggressive vs. Passive) X 2 (Group: Ingroup vs. Outgroup) repeated measures ANOVA on participants' likelihood to report the status to Facebook. As expected, we found a main effect for confrontation type, $F(1, 59) = 21.74$, $p < .001$, $\eta_p{}^2 = .27$. Participants were more likely to report statements that were accompanied by aggressive confrontations ($M = .59$, $SD = .40$) as compared to those with passive confrontations ($M = .45$, $SD = .40$). Similar to our findings for offensiveness, we also found a main effect for group membership, $F(1, 59) = 5.99$, $p = .02$, $\eta_p{}^2 = .09$. Statements accompanied with outgroup confronters ($M = .53$, $SD = .39$) were more likely to be reported than those with ingroup confronters ($M = .48$, $SD = .40$). There was no significant interaction between confrontation type and group membership, $F(1, 59) = .57$, $p = .45$.

Strength of Identity

To test H3 we conducted a hierarchical linear regression on our dependent variable of offensiveness with the same procedure as in Study 1. Our first model accounted for 8.56% of variance, and we found that confrontation type was significantly related to ratings of offensiveness, $\beta = .33$, $p = .04$, such that aggressive confrontations were related to higher ratings of offensiveness. Replicating Study 1, strength of identity was also related to offensiveness ratings, such that those who reported greater strength of identity, rated statements as more offensive, $\beta = .32$, $p < .001$. Our second and third model did not significantly account for any more variance, as compared to the first model, *ps* > .74, and none of our additional predictors significantly related to offensiveness ratings.

Discussion

We find consistency in our two main effects of confrontation and group membership on perceived offensiveness across both of our analyses. Aggressive and outgroup confronters was related to greater perceived offensiveness. Additionally, in both analyses we find that confrontation type was also significantly related to likelihood to report, such that aggressive confrontations led to more reporting of the post. However, only in our pre-registered analyses did group membership relate to likelihood to report. Similar to Study 1, we find that strength of identity did relate to perceived offensiveness, but this was not replicated in our multilevel analyses that was reported in the main text.

Study 3: Pre-registered analyses

We pre-registered this study (http://aspredicted.org/blind.php?x=7zz7ev) with hypotheses based upon our findings in both Study 1 and 2. H1& H2: Replicating Study 1 and 2, we anticipated that racist statements will be perceived as more offensive when presented with an aggressive confrontation. Given Study 2's results, we hypothesized that racist statements would be perceived as more offensive when presented with an outgroup confronter. Conversely, we hypothesized that attitudes towards the confronter would be more favorable to the ingroup vs. outgroup (H3). We hypothesized attitudes towards the confronter would be more positive when presented with an aggressive confrontation (H4). However, we expect that these two would interact, such that outgroup members would be perceived more favorably when presenting an aggressive confrontation (H5). H6 & H7: We hypothesized that those with a stronger sense of identity would rate statements as more offensive and have more positive attitudes towards ingroup confronters.

Results

Perceptions of Offensiveness

To test H1 and H2, we conducted a 2 (Confrontation: Aggressive vs. Passive) X 2 (Group: Ingroup vs. Outgroup) repeated measures ANOVA on ratings of offensiveness. As hypothesized, we found a main effect for confrontation type, $F(1, 60) = 29.71$, $p < .001$, $\eta_p^2 = .33$, such that statements accompanied by aggressive confrontations ($M = 5.88$, $SD = 1.34$) were rated as more offensive than those with passive confrontations ($M = 5.58$, $SD = 1.36$). We did not replicate our main effect for group membership seen in Study 2? Or MLM?, $F(1, 60) = 2.84$, $p = .10$. Additionally, there was no significant interaction between confrontation type and group membership, $F(1, 60) = .62$, $p = .43$.

Attitudes Towards the Confronter

To test H3-H5, we conducted a 2 (Confrontation: Aggressive vs. Passive) X 2 (Group: Ingroup vs. Outgroup) repeated measures ANOVA on attitudes towards the confronter. As hypothesized, there was a main effect for confrontation type, $F(1, 60) = 23.51$, $p < .001$, $\eta_p^2 = .28$, such that attitudes towards the confronter was more positive when the confronter was aggressive ($M = 4.29$, $SD = 1.37$), as compared to when they were passive ($M = 3.57$, $SD = 1.27$). As expected, we also found a main effect for group membership on attitudes towards the confronter, $F(1, 60) = 4.28$, $p = .04$, $\eta_p^2 = .07$, such that participants felt more positive towards ingroup confronters ($M = 4.01$, $SD = 1.19$) as compared to outgroup confronters ($M = 3.86$, $SD = 1.25$).

Lastly, as we anticipated, we found a confrontation type by group membership interaction for attitudes towards the confronter, $F(1, 60) = 9.40$, $p = .003$, $\eta_p^2 = .14$. Post-hoc comparisons using Tukey's correction were conducted (see Table 3). Confirming H5, when confrontations were aggressive, both ingroup and outgroup confronters were perceived positively. However, when confrontations were passive, participants viewed passive ingroup members more favorably as compared to passive outgroup members.

Table 3.

Mean ratings of attitudes towards the confronter by group and type of comment

Type of comment	Ingroup	Outgroup
Aggressive	4.26 (1.32)$_{c}$	4.32 (1.46)$_{ab}$
Passive	3.75 (1.32)$_{bcd}$	3.40 (1.39)$_{acd}$

Note: Standard deviations are in parentheses.
Paired subscripts indicate significant comparisons, *ps* < .01

Feeling Thermometers Towards the Confronter

To further examine positive or negative feelings towards the confronter, participants rated how warmly or coldly they felt toward the confronter with a Feeling Thermometer that ranged from *0=extremely cold* to *10=extremely warm*. This measure highly correlated with scale responses and therefore analyses not presented in main text.

We find similar results when looking at feeling thermometers towards the confronter, *F*(1, 60) = 22.60, $p < .001$, $\eta_p{}^2 = .27$. Participants felt warmer towards confronters who were aggressive (*M* = 7.26, *SD* = 2.12), as compared to passive (*M* = 6.02, *SD* = 2.17). However, the main effect for group membership was not significant, *F*(1, 60) = 3.50, *p* = .07.

We found an interaction between confrontation type and group membership on feeling thermometers, *F*(1, 60) = 6.15, *p* = .02, $\eta_p{}^2 = .09$. Post-hoc comparisons using Tukey's correction revealed the same pattern of results as the scale measure presented above: when confrontations were aggressive, both ingroup (*M* = 7.22, *SD* = 2.06) and outgroup (*M* = 7.30, *SD* = 2.31) confronters were perceived positively. However, when confrontations were passive, participants viewed ingroup members who were passive (*M* = 6.28, *SD* = 2.31; *p* = .01) significantly more favorably as compared to outgroup members who were passive (*M* = 5.76, *SD* = 2.27).

Strength of Identity

To test H6, we conducted a hierarchical linear regression on our dependent variable of offensiveness with the same procedures as outlined in Study 1. In the first model of our regression, we included all predictors: group, confrontation, and strength of identity. In the second model of our regression, we included all two-way interactions. In the last model of our regression, we included all three-way interactions. Our first model accounted for 15.00% of variance, and we found that confrontation type did not significantly relate to ratings of offensiveness, β= .31, *p* = .06. Replicating Study 1 and 2, strength of identity was also related to offensiveness ratings, such that those who reported greater strength of identity, rated statements as more offensive, β= .39, $p < .001$. Our second and third model did not significantly account for any more variance, as compared to the first model, *ps* > .86, and none of our additional predictors significantly related to offensiveness ratings.

To test H7, we conducted the same hierarchical linear regression on our dependent variable of attitudes towards the confronter. For our dependent variable of attitudes towards the confronter, our first model accounted for 6.55% of variance, and we found that confrontation type was significantly related to attitudes towards the confronter, β= .70, $p < .001$, such that aggressive confrontations were related to more positive attitudes towards the confronter. Interestingly, we did not find the relationship between strength of identity and attitudes towards the confronter, β= -.02, *p* = .79. Similarly, group membership was not related to attitudes towards the confronter, β= -.15, *p* = .40. Our second and third model did not significantly account for any more variance, as compared to the first model, *ps* > .38.

Discussion

Again, we find some discrepancies with our pre-registered analyses and multilevel analyses. Both analyses show a main effect of confrontation type, such that aggressive confrontations related to greater perceived offensiveness. Similarly, we see that more highly identified participants also perceived the post as more offensive. However, we do not replicate the effect of group membership in our pre-registered analyses. We do replicate our findings in both analyses for our dependent variable of attitudes towards the confronter. In both analyses, we find main effects for confrontation type and group membership, as well as a significant interaction, suggesting that when confrontations were aggressive, all confronters were seen positively, yet, when confrontations were passive outgroup members were seen less positively. We did not find any relationships between strength of identity and attitudes towards confronters in our pre-registered analyses.

Overall Summary

Overall, we find general consistencies across our pre-registered analyses and our multilevel analyses strategy. Our findings concerning confrontation type are robust across three studies and two analysis strategies. Aggressive confrontations relate to greater perceived offensiveness, increased likelihood to report a racist post, and more positive evaluations of the confronter. What is less certain is the role that group membership plays in Asian American participants' perceptions. While in Study 2, both analysis strategies revealed a main effect for group membership, we do not find consistency across Studies 1 and 3. Similarly, we find discrepancies in the relationship between strength of identity and perceptions of offensiveness. In our pre-registered analyses, these effects were consistent across studies, however, in our multilevel analyses these effects largely disappear until Study 3. Overall, our most robust findings across these two analytic strategies are confrontation type on all measured dependent variables and the interaction between confrontation type and group members on attitudes towards confronters.

One reason we may be observing slightly different pattern of results for group membership across the two analysis strategies is how multilevel analyses is better able to capture individual differences and variability across stimuli trials. By not aggregating participants' responses across all 16 stimuli we may be more likely to control for random variability across each stimuli. Within the field, it is conventional to analyze repeated measures designs with a repeated ANOVA, which is why we pre-registered this as our analysis strategies. However, there are clear advantages to using a multilevel analytic approach, such as reduced Type I error, issues with multiple comparisons, and controlling for random effects within stimuli (see Gelman et al., 2012; Judd et al., 2012). For these reasons we report multilevel analyses in main text, but for transparency we have included our full pre-registered analyses.

Study 2: Exploratory Analyses

Methods

Likelihood to Respond

Below the statement stimuli, participants were provided an empty text box with the prompt "Write a comment…" similar to the one that is seen on Facebook's comment box. Participants had the option of not responding, simply by proceeding to the next page.

Content of Responses

We coded responses in two ways. First, for the presence of a comment (coded as 0 for no response, and 1 for typed response). Second the content of the responses were coded for confrontation valence (passive-aggressive). Two research assistants were trained to rate all the open responses on these two dimensions. To capture confrontation valence, we had responses rated for how passive-aggressive they were on a scale of *1 = very passive* to *7 = very aggressive*. The ratings from the two coders were averaged into a final score for each comment (α =.70).

Results

We used a multilevel analytic strategic, as outlined in Study 1.

Exploratory Analyses: Likelihood to Respond

We examined whether Confrontation, Group, and Identity had an impact on participants' likelihood to respond. We found no main effects for confrontation type, *b* = -2.18, *SE* = 2.54, 95% CI [-7.16, 2.81], exp(B) = .11, *z* = -.86, *p* = .39. There was also no main effect for group membership, *b* = -.66, *SE* = 2.11, 95% CI [-4.79, 3.47], exp(B) = .52, *z* = -.31, *p* = .75. There was no interaction between confrontation type and group membership, *b* = -.73, *SE* = .77, 95% CI [-2.24, .78], exp(B) = .48, *z* = -.95, *p* = .34. Strength of identity and its interaction terms were not significant, *p*s > .32.

Exploratory Analyses: Content of Responses

We also examined whether the content of responses were influenced by Confrontation, Group, and Identity[5]. We found no main effect for confrontation type, *b* = -.18, *SE* = 16, 95% CI [-.50, .14], *t* = -1.11, *p* = .27. However, we did find a main effect for group membership, *b* = -.62, *SE* = .17, 95% CI [-.94, -.29], *t* = -3.74, *p* < .001. Participants' responses were more aggressive when shown an outgroup confronter (*M* = 3.86, *SD* = .79), as compared to an ingroup confronter (*M* = 3.36, *SD* = .90). There was no significant interaction, *b* = .35, *SE* = .23, 95% CI [-.09, .80], *t* = 1.56, *p* = .12. Strength of identity, and all 2-way and 3-way interactions were not significant, *p*s > .51.

Study 3: Exploratory Analyses

Methods

Likelihood to Respond

Same procedure as Study 2.

Content of Responses

As outlined in Study 2 above, responses were coded for how passive-aggressive they were on a scale of *1 = very passive* to *7 = very aggressive*. The two ratings were averaged into a final score for each comment (α = .75).

Results

We used a multilevel analytic strategic, as outlined in Study 1.

Exploratory Analyses: Likelihood to Respond

We examined whether Confrontation, Group, and Identity had an impact on participants' likelihood to respond. We found no main effects for confrontation type, *b* = 1.25, *SE* = .71, 95% CI [-.13, 2.63], exp(B) = 3.49, *z* = 1.77, *p* = .08. We did however, find a main effect for group membership, *b* = 2.75, *SE* = 1.06, 95% CI [.68, 4.82], exp(B) = 15.61, *z* = 2.60, *p* = .009. Statements that were accompanied with ingroup confronters were more likely to elicit responses (*M* = 73%, *SD* = 41%) than those with outgroup confronters (*M* = 70%, *SD* = 41%). In other words, participants were approximately

[5] Only 27 participants responded to posts, thus these analyses were conducted on *n* = 27.

15.61 times more likely to respond when the post included an Asian confronter as compared to a White confronter. There was no interaction between confrontation type and group membership, b = -.49, SE = .70, 95% CI [-1.87, .88], exp(B) = .61, z = -.70, p = .48. Strength of identity and its interaction terms were not significant, ps > .25.

Exploratory Analyses: Content of Responses

We also examined whether the content of participants' responses was influenced by Confrontation, Group, and Identity[6]. We found no effect for confrontation type, b = .06, SE = .12, 95% CI [-.17, .30], t = .52, p = .60. Similar to Study 2, we found a main effect for group membership, b = -.25, SE = .12, 95% CI [-.50, -.01], t = -2.05, p = .04. Statements accompanied with an outgroup confronter elicited more aggressive responses (M = 3.37, SD = 1.00) as compared to those with ingroup confronters (M = 2.97, SD = 1.06). However, unlike Study 2, this was qualified by a significant interaction, b = .64, SE = .17, 95% CI [.30, .97], t = 3.74, p < .001. Simple effects analyses with group membership as a moderator showed that for regardless whether confrontation was aggressive or passive, posts with outgroup confronters elicited similar responses from participants, b = .06, SE = .12, 95% CI [-.18, .30], t(459) = .52, p = .60. However, posts that included aggressive ingroup confronters elicited more aggressive responses as compared to passive ingroup, b = .70, SE = .12, 95% CI [.46, .94], t(459) = 5.81, p < .001. We also analyzed simple effects with confrontation type as the moderator. We found that when confrontation was passive there was a significant difference across group membership, such that outgroup confronters elicited more aggressive responses than ingroup confronters, b = -.25, SE = .12, 95% CI [-.50, -.009], t(174) = -2.05, p = .04. On the contrary, when confrontations were aggressive: ingroup confronters elicited significantly more aggressive responses than outgroup confronters, b = .38, SE = .12, 95% CI [.14, .63], t(179) = 3.08, p = .002. Tentatively, we suggest that our Asian Participants were more likely to mirror the responses made by ingroup members; if an ingroup confronter was aggressive, the responded aggressively as well and vice versa. However, if a confronter was an outgroup member, confrontation

[6] Only 45 participants responded to posts, thus these analyses were conducted on n = 45.

type did not impact participants' responses. Strength of identity and its interaction terms was not significantly related to content of responses, *p*s > .16.

Overall Summary

Overall, we find in our exploratory measures that group membership of confronter matters, however, our preliminary results are not consistent. Asian American participants are more likely to respond to outgroup confrontations, which falls into line with group norm theory. This suggest that if the majority group responds to a racist post, others will follow this behavior and respond similarly. In Study 3, while we find that participants are still more likely to respond when seeing outgroup members confront, they are more likely to tune the valence of their response with ingroup confronters, which suggest that they may be social referencing other ingroup members for the appropriate type of response to make. This might suggest that there are different mechanisms to explain the different stages of intergroup interactions online. From interpreting a racist event, motivation to react and respond to the event, and strategy in those confrontations. Online interactions are dynamic and perceptions are influenced by other's in this space, however more research needed into how intergroup dynamic come into play in this process.

Study 3: Feeling Thermometers

Results

Similar to our findings with our attitude measure there was a main effect for Confrontation, *p* < .001, such that attitudes towards the confronter was more positive when the confronter was aggressive (vs. passive). As expected, we also found a main effect for Group on attitudes towards the confronter, *p* < .001, such that participants felt more positive towards ingroup confronters as compared to outgroup confronters. Lastly, as anticipated, we found a Confrontation X Group interaction for attitudes towards the confronter, *p* = .006. Simple effects analyses found that when confrontations were aggressive, both ingroup and outgroup confronters were perceived positively, *b* = -.06, *SE* = .15, 95% CI [-.36, .24], *t*(203) = -.40, *p* = .69. However, when confrontations were passive, participants viewed ingroup members who were passive significantly more favorably as compared to outgroup members who were passive, *b* = .52, *SE* = .15, 95% CI [.22, .82], *t*(203) = 3.38, *p* <.001. In addition, we found an interaction between Group X Identity, *p* = .02. Simple effects analyses show that for participants who were weak (-1 *SD*) in Asian identification did not differ their attitudes towards ingroup vs. outgroup confronters, *b* = -.07, *SE* = .16, 95% CI [-.38, .25], *t*(60) = -.44, *p* = .66. However, participants who had strong Asian identities (+1 *SD*) viewed ingroup confronters more favorably than outgroup confronters, *b* = .53, *SE* = .16, 95% CI [.21, .84], *t*(60) = 3.34, *p* = .001. See Table 4 for parameter estimates.

Table 4.

Parameter estimates for feeling thermometers in Study 3

Effect	Variance	SD	Estimate	*SE*	95% CI	*t*	*df*	*p*
Random effects								
Participants								
Intercept	4.64	2.16						
Confrontation	3.31	1.82						
Group	.08	.29						
Stimuli								
Intercept	.020	.14						

Residual	2.73	1.65						
Fixed effects								
Intercept			5.77	.30	5.17, 6.36	19.16	61.40	<.001
Confrontation (C)			1.51	.28	.97, 2.05	5.51	82.10	<.001
Group (G)			.52	.15	.22, .82	3.38	202.90	<.001
Identity (I)			-.26	.23	-.71, .18	-1.16	64.00	.25
C X G			-.58	.21	-.99, -.16	-2.75	801.00	.006
C X I			.35	.21	-.07, .77	1.64	82.10	.10
G X I			.28	.12	.05, .51	2.37	202.90	.02
C X G X I			-.10	.16	-.42, .22	-.62	801.00	.53