

RESEARCH ARTICLE

Ally confrontations as identity-safety cues for marginalized individuals

Laura K. Hildebrand¹  | Celine C. Jusuf² | Margo J. Monteith¹

¹Purdue University, West Lafayette, IN, USA

²San Diego State University, San Diego, CA, USA

Correspondence

Laura K. Hildebrand, Psychological Sciences, Purdue University, 703 3rd Street, West Lafayette, IN 47907, USA.

Email: HildebrandLK@gmail.com

Funding information

National Science Foundation, Grant/Award Number: DGE-1333468; Purdue Acceptance and Inclusion Consortium

Abstract

Three studies and an integrative data analysis ($N = 1,017$) demonstrated that confrontations (speaking up against a stereotypical or prejudiced statement), when affirmed by bystanders, serve as an effective safety cue for targets of bias. In Studies 1 and 2, Chinese-American and White women witnessed anti-Asian and sexist remarks, respectively. Results revealed that a lone confronter (i.e., a confronter not affirmed by others) was unable to boost identity-safety (e.g., belonging, safety) compared to when the bias was not confronted, regardless of confronter identity (i.e., ally vs. in-group confronter). Study 2 demonstrated that other people in the interaction group (i.e., bystanders) must affirm the confrontation for it to serve as an effective safety cue. Study 3 replicated and extended these results among White women for confrontation of sexism and Black women for confrontation of racism. Overall, these studies suggest that confrontations, when affirmed, can serve as a safety cue.

KEYWORDS

belonging, bias, confrontation, prejudice, safety cues, stereotyping

1 | INTRODUCTION

In a 2016 Olympic interview, the interviewer incorrectly described Andy Murray as “the first person ever to win two Olympic tennis gold medals”. Murray, a White man, immediately corrected the interviewer, stating that Venus and Serena Williams, two Black women, “have won about four [gold medals] each” (Barr, 2019). Even though the interviewer’s mistake did not personally affect him, Murray did not ignore it. Instead, he confronted. One can imagine how the Williams sisters, as well as other Black women, may have felt as a result of the interviewer’s statement. Perhaps they felt invisible in the Olympic environment, or like they didn’t belong. Could Murray’s confrontation make the Williams sisters feel safer and more welcome?

Or do biased expressions create a more indelible threat to marginalized group members’ feelings of belonging and safety?

Allies are individuals who, although not a member of the group targeted by bias, nonetheless take action to reduce discrimination and increase social justice (Becker & Barreto, 2019; Drury & Kaiser, 2014). In this example, Murray acted as an ally by choosing to confront the biased statement. The present research investigated whether ally confrontation can restore target group members’ feelings of safety and belonging in situations where biased comments have occurred. Do ally confrontations increase targets’ sense of belonging and safety? How are ally confrontations perceived compared to confrontations by ingroup members or members of other disadvantaged groups? How important are the reactions of bystanders, and do they need to affirm confrontations before targets’ feelings of safety and belonging will increase? We examine these questions across three experiments by manipulating confrontation, confronter identity, and bystander behavior while examining targets’ reactions to the expression of anti-Asian, anti-Black, and anti-woman stereotypical comments.

Hildebrand and Jusuf contributed equally and are listed alphabetically.

This research was conducted in compliance with the ethical principles outlined in the Belmont Report (Respect for Persons, Beneficence, and Justice). All studies were approved by the Purdue Human Research Protection Program and its associated Institutional Review Board.

1.1 | Confrontation of bias

Although social norms often condemn bias (such as racial or gender stereotypes), biased remarks are still common in everyday conversations (e.g., Swim & Hyers, 1999; Swim, Hyers, Cohen, Fitzgerald, & Bylsma, 2003). These remarks may reduce the target individual's desire for interaction with people involved in the conversations and threaten their sense of belonging (Richman & Leary, 2009). However, to combat such effects, one can speak up, or confront, these biased remarks.

Research suggests that confrontation is a powerful tool for curbing people's expressions of bias. After confrontation, people are less likely to express bias both immediately and several days later (Burns & Monteith, 2019; Chaney & Sanchez, 2018a; Czopp, Monteith, & Mark, 2006; Gulker, Mark, & Monteith, 2013; Mallet & Wagner, 2011; Parker, Monteith, Moss-Racusin, & Van Camp, 2018; Rasinski & Czopp, 2010; for a review, see Mallet & Monteith, 2019). However, a robust finding in this literature is that who does the confronting (i.e., confronter identity) influences confrontation effectiveness. Specifically, confrontation by a dominant group ally (e.g., a man confronting sexism, or a White person confronting anti-Black prejudice) produces greater bias reduction than confrontation by a target group member (e.g., a woman confronting sexism, or a Black person confronting anti-Black prejudice; Czopp & Monteith, 2003; Gervais & Hillard, 2014; Gulker et al., 2013; Rasinski & Czopp, 2010; Schultz & Maddox, 2013). This differential effectiveness appears to stem from perceptions of target group confronters as complainers and overly sensitive, whereas dominant group ally confrontations are perceived as more legitimate and appropriate (Czopp et al., 2006; Eliezer & Major, 2012; Kaiser, Hagiwara, Malahy, & Wilkins, 2009; Rasinski & Czopp, 2010).

How do target group members perceive confrontations on their behalf? Target group members often perceive allies as interpersonally supportive (Brown, 2015; Brown & Ostrove, 2013), especially to the extent that they provide autonomy-oriented, rather than dependency-oriented, help (i.e., supporting disadvantaged group members in their efforts to address bias rather than taking over these efforts; Wiley & Dunne, 2019; see also Droogendyk, Wright, Lubensky, & Louis, 2016). Recent research examining ally confrontation revealed that targets perceive allies who confront bias as acting appropriately and as not overreacting, and sometimes more so than target confronters (Kutlaca, Becker, & Radke, 2020). Psychological and performance outcomes may also be enhanced when allies confront bias on targets' behalf. For instance, in Cihangir, Barreto, and Ellemers's (2014) research, women participants were interviewed by a male who exhibited subtle sexism through questions he asked. Next, other men or women who had also completed the interview questions acknowledged and censured the sexism, or it was left unacknowledged. On subsequent tasks, the women participants appeared to suffer on measures of self-confidence and performance, but not if men had confronted the sexism in the interview. Thus, ally confrontation appeared to alleviate bias-induced social identity threat.

1.2 | Confrontation as an identity-safe cue for marginalized-group members

Given findings that confrontation can reduce the expression of bias and alleviate social identity threat, we reasoned that it might also signal to targets that their identities will be safe in environmental contexts (e.g., social, work, and educational settings) where bias has been expressed. Bias threatens a marginalized group member's identity, which in turn may lead to decreased feelings of belonging and interest within the relevant environment (Branscombe, Schmitt, & Harvey, 1999; Goffman, 1963; Murphy, Steele, & Gross, 2007). Confrontation, however, may act as a safety cue that counteracts this threat. Safety cues are environmental cues (e.g., organizational diversity structures, gender-inclusive bathrooms) that signal a person's identity is valued and promote belonging and trust (i.e., identity-safety; Chaney & Sanchez, 2018b; Cheryan, Plaut, Davies, & Steele, 2009; Wout, Murphy, & Barnett, 2014). Confrontations may similarly signal identity valuation and promote identity-safety because they often directly contradict a prejudiced or stereotypic remark and communicate that the confronter values the targeted identity and an egalitarian norm (Shelton, Richeson, Salvatore, & Hill, 2006).

1.3 | Confronter identity

The extent to which confrontation serves as a safety cue may depend on who does the confronting. We considered three types of possible confronters: a confronter from the target group, an ally who is a member of another similarly marginalized group, and an ally who belongs to the dominant group with respect to the dimension of identity being targeted. For instance, imagine that an Asian person observing a social interaction hears someone make an anti-Asian comment that is then confronted by another person in the interaction. Would the observing Asian person experience greater identity-safety when the confrontation comes from an Asian person in the interaction group (ingroup confronter), from a Black person (other marginalized group ally), or from a White person (dominant group ally)?

An ingroup member's confrontation and, perhaps to a lesser extent, a confrontation by a marginalized group ally, may signal solidarity (Craig & Richeson, 2012; 2016) and help to establish a sense of identity-safety within the given environment. Members of targeted groups are most likely to expect and find support from their own ingroup (Simon & Klandermans, 2001; Swim & Hyers, 1999; Tajfel & Turner, 1986; Warner & Branscombe, 2012). Marginalized allies may also provide a unique inclusive identity that supports solidarity through similar histories of oppression and mutual benefit finding (Affleck & Tennen, 1996; Craig & Richeson, 2012; Warner, Wohl, & Branscombe, 2014). Additionally, members of stigmatized groups believe that, compared to dominant group allies, both ingroup members and similarly marginalized allies have a greater awareness of social injustice and inequality (Brown & Ostrove, 2013). This sense of

solidarity may enable ingroup and marginalized ally confrontations to act as effective safety cues.

On the other hand, targets of bias may perceive dominant group confrontation as a stronger signal that bias will not be condoned in a given environment. As we summarized above, dominant group ally confronters appear to curb bias more effectively than other confronters, often because they are not as likely to be dismissed as complainers or as seeing injustice where it does not exist (Czopp & Monteith, 2003; Drury & Kaiser, 2014; Eliezer & Major, 2012; Gervais & Hillard, 2014; Kaiser & Miller, 2001; Kutlaca et al., 2020; Schultz & Maddox, 2013). Targets may be aware of this powerful influence of dominant group members over others (Droogendyk et al., 2016). As a result, dominant group allies may provide a stronger signal that one's identity and safety will be protected in environmental contexts involving interactions with others. Furthermore, dominant group allies' confrontations may provide more potent validation of one's social identity than confrontations by ingroup members (Cihangir et al., 2014) or marginalized group allies.

1.4 | Group norms and affirmed confrontations

We also considered the possibility that, when targets consider their identity-safety in an environment after bias has been expressed, the behavior of other people who are also present (i.e., bystanders) will be important. That is, identity-safety may hinge not just on one person speaking out against bias, but also on bystanders joining in to affirm the confrontation. Even if one person has stepped up to express discontent with a biased remark, if other people present in the situation do not affirm that confrontation, targets may continue to doubt that their social identities will be valued and safe in the environmental context. For instance, the silence of confrontation bystanders may communicate that they fear the person who made the biased statement (Swim & Hyers, 1999; Woodzicka & LaFrance, 2001), or that they agree with or simply do not care about the expression of bias. If this is the case, one person's confrontation, regardless of their identity, may be insufficient for a target group member to infer identity-safety.

However, if other people in that environment express support for the confrontation, targets of bias should be more likely to perceive norms that do not tolerate bias. More specifically, affirmations of the confrontation by bystanders should establish both injunctive and descriptive norms (Cialdini, Reno, & Kallgren, 1990; Reno, Cialdini, & Kallgren, 1993), thereby communicating that bias is neither condoned nor characteristic of the environmental context. Such norms cause people to adjust their behavior accordingly (Ata, Bastian, & Lusher, 2009; Miller, Monin, & Prentice, 2000; Smith & Louis, 2008), and reduce expressions of prejudice (Allport, 1954; Crandall, Eshleman, & O'Brien, 2002; Crandall & Stangor, 2005). Therefore, when a confrontation is supported by other people who are also present, relative to when no support is provided, targets of bias may feel more confident that their social identities will not be devalued in the setting.

1.5 | Overview of the present research

Across three studies, we investigated whether confrontation acts as a safety cue for targets in a setting where a biased comment was made. Specifically, would targets report greater belonging, safety, and desire to join the interaction group (i.e., identity-safety) when a biased comment was confronted (Studies 1–3)? In addition, we assessed evaluation of the interaction group (i.e., how warmly, or positively, they felt toward the group) to gauge overall explicit attitude toward the group (e.g., Nosek, 2007). Would the identity of the confronter (i.e., target group member, other marginalized group ally, dominant group ally) play a role (Studies 1 and 2)? Must non-confronting group members affirm a confrontation for it to act as a safety cue (Studies 2 and 3)?

In Study 1, we examined two competing hypotheses concerning confronter identity. Specifically, we considered it theoretically possible that either (a) dominant group ally confrontations (i.e., a White confronter) or (b) ingroup confrontations (i.e., an Asian confronter) might provide a better safety cue and elicit more positive feelings toward the interaction group following an anti-Asian remark. In Study 2, we tested these same competing hypotheses within the context of gender bias. We also examined the role of confrontation affirmations by non-confronting interaction group members, with the hypothesis that affirmations would increase targets' warmth toward the group and identity-safety. Finally, in Study 3, we tested competing hypotheses that affirmed confrontations would enhance targets' positive feelings and identity-safety (a) equally in the context of gender and racial bias, versus (b) more for gender than racial bias. In all experiments, we controlled for individual differences in sociability and desire for group interactions (e.g., Wilt & Revelle, 2017) by using ratings participants completed for a filler group interaction when predicting dependent variables for the critical group interaction.

These hypotheses were explored among three participant populations (i.e., Qualtrics-recruited Chinese-American women, undergraduate participants, and MTurk-recruited Black and White women) and three types of bias (anti-Asian bias, anti-women bias, and anti-Black bias).

2 | STUDY 1

2.1 | Method

2.1.1 | Participants and design

Given the novelty of the present research questions, we did not have directly relevant prior effect sizes to inform a power analysis. We considered that (a) most effects in social/personality psychology range from small to medium (Richard, Bond, & Stokes-Zoota, 2003), (b) confrontation research routinely produces medium-sized effects (e.g., Czopp et al., 2006), and (c) our sample was difficult to recruit. Given these considerations, we conducted a power analysis with G*Power v.3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007) using a medium-sized effect ($f = 0.25$), which indicated that 200 participants

(or, approximately 40 participants per cell) would provide 80% power to detect this effect size.

We recruited 214 Chinese-American women between the ages 25 and 40 ($M_{\text{age}} = 30.13$, $SD_{\text{age}} = 3.25$) through an online Qualtrics Panel.¹ Age was restricted to increase the likelihood that participants could relate to the critical script content. Each participant received \$4.95. Participants were randomly assigned to one of five conditions in a single-factor between-participants design. Three conditions included an anti-Asian remark that was confronted by either a Chinese-American woman, a Black woman, or a White woman. In a fourth condition, the anti-Asian remark was not confronted. In a fifth condition, the anti-Asian remark was not made.

2.1.2 | Procedure

After consenting to participate, participants learned that the study examined perceptions of group interactions. Participants first read a filler script, which was included to mask the study's focus. Before the script started, participants read a short description of each character, accompanied by their name and picture. In the script, each character line identified the person "speaking" by name and picture so participants would know who said what. Afterwards, participants completed all measures (see below) in relation to this group.

Participants were randomly assigned to receive one of several versions of the second, critical script. The context was described as four women meeting for coffee after dropping their children off for the first day of school. Participants viewed a short description of each character, including names and pictures to convey each person's race (two White women, one Chinese-American woman, one Black woman). Then, participants read a script of the interaction, again with each line accompanied by the person's picture and name.

In the conditions including a biased remark, one of the White women stated, "Asians aren't really good drivers typically, so I'm just worried...I'm not sure I'd be comfortable with this [carpooling] arrangement". In conditions involving a confrontation, the remark was confronted with, "Hey, that's just a stereotype. Some Asians are bad drivers, some aren't. Just like other people. You can't assume [she's/I'm] going to be a bad driver just because [she's/I'm] Asian!" The confronter identity was manipulated to be either the other White woman, the Black woman, or the Chinese-American woman. When the confrontation occurred, the script ended immediately after it. In the No Confrontation condition, the script ended with the unopposed anti-Asian remark. In the No Comment condition, no biased remark was made.

Then, the full study purpose was disclosed, and participants completed a post-session consent form that asked for permission to use their data.

¹Per our inclusion criteria, the Qualtrics Panel (see <https://www.qualtrics.com/uk/research-services/online-sample/>, for more on Qualtrics Panels) only provided data for participants who passed two out of three attention checks embedded in the survey, took at least 12 min to take the survey, and responded correctly to two items that asked about critical script information.

2.1.3 | Measures²

Perpetrator Offensiveness

As a check that the perpetrator's (i.e., person who made the biased remark) remark was perceived as offensive, participants rated how rude, offensive, and ignorant this character was on a 1 (*Strongly Agree*) to 7 (*Strongly Disagree*) scale. Items were averaged to create a composite *offensiveness* score.

Feeling toward the Interaction Group

Participants reported their feeling toward the interaction group (i.e., the women who interacted in the script) using a feeling thermometer ranging from 0 (*cold*) to 100 (*warm*).

Identity-safety

Based on prior literature (e.g., Cheryan et al., 2009), identity-safety was assessed with measures of *belonging* (adapted from Walton & Cohen, 2007; seven items; e.g., "Other members of this group would accept me"), *safety* (adapted from Purdie-Vaughns, Steele, Davies, Dittmann, & Crosby, 2008; Good, Rattan, & Dweck, 2012; seven items; e.g., "I think that I could trust this group to treat me fairly"), and *desire to join the group* (adapted from Purdie-Vaughns et al., 2008; seven items; e.g., "I think it would be pleasant to be a part of this group") on a 1 (*Strongly Disagree*) to 7 (*Strongly Agree*) scale. Presentation of the items was randomized.

2.2 | Results

2.2.1 | Analytic approach

We observed high correlations between the identity-safety measures ($r_s > .74$), which also proved to be the case for subsequent studies. Therefore, we conducted confirmatory factor analyses (CFA) using MPlus (Version 8.0, Muthén & Muthén, 2017), including data from all studies, to test whether a one-factor model fit the data better than a three-factor model (see Figure 1). The three-factor model provided an excellent fit (according to Hu & Bentler's, 1999, criteria); however, most fit indices indicated the one-factor model was acceptable (see Table 1). Further, results were largely redundant across the identity-safety measures when analyzed separately (see Supplemental Materials). Thus, for all studies, results based on the one-factor measure, *identity-safety*, are presented.

Perpetrator Offensiveness was predicted using condition (No Confrontation, Chinese-American Confronter, White Confronter, Black Confronter, and No Comment) in a between-participant analysis of variance (ANOVA). For analyses of Feeling toward the Interaction Group

²Participants also completed two exploratory measures: (a) identification as an Asian woman, and (b) confronter perceptions (i.e., the extent to which the confronter was seen as a complainer and hypersensitive). However, condition did not influence either variable, and so we do not discuss them further. See Appendix S1 for details.

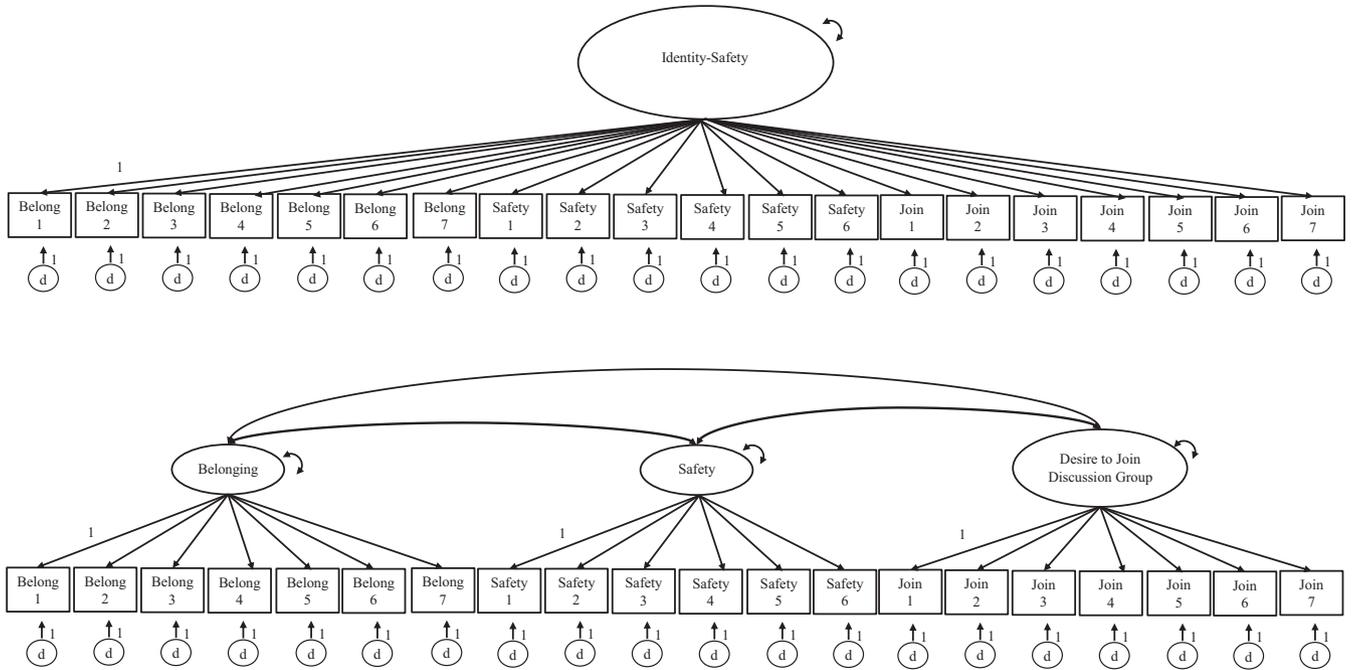


FIGURE 1 Models tested in confirmatory factor analyses

TABLE 1 Fit statistics for CFA models, Study 1

	χ^2	df	RMSEA	RMSEA 90%CI	CFI	TLI	SRMR	AIC	BIC
3-Factor Model	1,262.45*	167	0.08	<0.0001	0.96	0.95	0.03	61,387.71	61,703.36
1-Factor Model	2,573.755*	170	0.113	<0.0001	0.90	0.89	0.04	62,693.01	62,993.63

* $p < .00001$.

and Identity-safety, we also included a covariate (thus, performing ANCOVAs). Based on an a priori decision, participants' scores on the corresponding measure from the filler script were used as a covariate in feeling and identity-safety analyses for this and all subsequent studies to control for these individual differences. Their inclusion affected significance in only one instance, which we have footnoted where the relevant result is presented. See Supplemental Materials for all analyses without covariates (i.e., ANOVAs).

For this and all subsequent studies, follow-up comparisons were performed with Bonferroni corrections (i.e., $\alpha/10$ for Study 1, $\alpha/6$ for Studies 2 and 3), and all reported ps reflect these corrections. We report confidence intervals for the difference between relevant means involved in comparisons. Small differences in degrees of freedom sometimes occur due to missing data.

See Table 2 for descriptive statistics, reliability, and inter-measure correlations and Table 3 for cell means and standard deviations.

2.2.2 | Perpetrator offensiveness

The ANOVA revealed that an effect of condition on offensiveness ratings was significant, $F(4, 209) = 21.50, p < .001, \eta_p^2 = .29$. Verifying

that the biased comment was interpreted as intended, the perpetrator was seen as more offensive in conditions where a biased comment was made compared to the No Comment condition, regardless of whether the confronter was Chinese-American, $t(209) = 7.47, SE = 0.33, p < .001, 95\%CI [-3.41, -1.52]$, Black, $t(209) = 7.93, SE = 0.34, p < .0001, 95\%CI [-3.61, -1.35]$, White, $t(209) = 7.39, SE = 0.33, p < .000, 95\%CI [-3.41, -1.52]$, or no confrontation was made, $t(209) = 6.53, SE = 0.37, p < .001, 95\%CI [-3.42, -1.35]$. No other comparisons were significant, $ps = 1.00$.³

2.2.3 | Feeling toward interaction group

Feelings toward the filler interaction group (i.e., the group in the filler script) was a significant covariate when predicting feelings toward the critical interaction group, $F(1, 207) = 20.59, p < .001, \eta_p^2 = .09$. Of greater interest, the ANCOVA revealed a significant main effect for condition, $F(4, 207) = 4.53, p = .002, \eta_p^2 = .08$. Participants felt warmer toward the interaction group when no biased comment was made than when it was made but not confronted, $t(207) = 3.46,$

³Values of $p = 1.00$ often emerged due to the Bonferroni corrections applied.

TABLE 2 Descriptive statistics, reliability, and correlations between measures, Study 1

	M (SD)	α	1.	2.
1. Perpetrator Offensiveness	4.78 (1.83)	.91	-	
2. Feeling toward Interaction Group	62.91 (23.62)	-	-.47*	-
3. Identity-safety	4.15 (1.28)	.97	-.43*	.78*

Note.: Note: Feeling toward the interaction group was a one-item measure, so no alpha is reported.

* $p \leq .001$.

$SE = 5.16$, $p = .007$, 95%CI [-32.53, -3.23]. However, confrontation did not increase warmth. The No Confrontation condition did not differ from conditions with a Chinese-American confronter, $t(207) = 0.30$, $SE = 4.99$, $p = 1.00$, 95%CI [-15.71, 12.66], Black confronter, $t(207) = 0.22$, $SE = 5.01$, $p = 1.00$, 95%CI [-15.33, 13.10], or White confronter, $t(207) = 1.02$, $SE = 4.99$, $p = 1.00$, 95%CI [-19.25, 9.04]. Finally, greater warmth was reported in the No Comment condition than with the Chinese-American confronter, $t(207) = 3.46$, $SE = 4.73$, $p = .007$, 95%CI [2.95, 29.77], and the Black confronter, $t(207) = 3.52$, $SE = 4.75$, $p = .005$, 95%CI [3.30, 30.23]. Providing some, albeit weak, support that the White confronter served as a safety cue, warmth was comparable in the No Comment and White Confrontation conditions, $t(207) = 2.70$, $SE = 4.72$, $p = .07$, 95%CI [-0.62, 26.17].

2.2.4 | Identity-safety

Filler identity-safety was a significant covariate, $F(1, 208) = 13.64$, $p < .001$, $\eta_p^2 = .06$. More importantly, the ANCOVA yielded a significant main effect for condition, $F(4, 208) = 4.90$, $p = .001$, $\eta_p^2 = .09$. As with feeling toward the interaction group, participants felt more identity-safety when no biased comment was made than when a biased comment was made but not confronted, $t(208) = 3.60$, $SE = 0.28$, $p = .004$, 95%CI [-1.80, -0.21]. Furthermore, the No Confrontation condition again did not differ from conditions with a Chinese-American confronter, $t(208) = 0.06$, $SE = 0.27$, $p = 1.00$, 95%CI [-0.79, 0.75], Black confronter, $t(208) = 0.61$, $SE = 0.27$, $p = 1.00$, 95%CI [-0.94, 0.61], or White confronter, $t(208) = 1.26$,

$SE = 0.27$, $p = 1.00$, 95%CI [-1.11, 0.43]. Finally, greater identity-safety was reported with no comment than with the Chinese-American confronter, $t(208) = 3.87$, $SE = 0.26$, $p = .001$, 95%CI [0.26, 1.71], and the Black confronter, $t(208) = 3.27$, $SE = 0.26$, $p = .01$, 95%CI [0.11, 1.56]. Again, providing some, albeit weak, support for the idea that the White confronter could serve as a safety cue, identity-safety was comparable in the No Comment and White Confrontation conditions, $t(208) = 2.59$, $SE = 0.26$, $p = .01$, 95%CI [-0.06, 1.39].

2.3 | Summary

An anti-Asian statement caused Chinese-American participants to report colder feelings toward an interaction group and less identity-safety (belonging, safety, and desire to join the group). Furthermore, confrontation did little to counteract this effect: When the biased comment was confronted, participants did not report greater identity-safety than when the biased comment was not confronted. The White confronter appeared to boost identity-safety somewhat, but in no case were levels of identity-safety greater when the White woman confronted the anti-Asian comment than when the comment was not confronted.

3 | STUDY 2

Study 2 further explored confrontation as a safety cue, this time in the context of a sexist comment that was or was not confronted by a woman (ingroup) or a man (ally). Given there was no significant difference between any of the confrontation conditions and the No Confrontation condition in Study 1, we sought to increase power to detect significant effects if they were indeed to be found. In addition to increasing sample size, we strengthened the manipulation by having participants listen to an audio recording of the interaction, rather than reading a script. Perhaps the biased statement was especially salient when it was read in Study 1, thereby limiting attention to subsequent content, including the confrontation. Also, reading the interaction may not have been very engaging, again potentially dampening confrontation effects. Finally, the ingroup member's confrontation in Study 1 involved the same person toward whom

TABLE 3 Measures as a function of confrontation condition, Study 1

Measure	No Confrontation $n = 33$	Black Confronter $n = 46$	White Confronter $n = 47$	Chinese Confronter $n = 47$	No Comment $n = 41$
Perpetrator Offensiveness	5.14 _a (1.99)	5.41 _a (1.44)	5.22 _a (1.64)	5.22 _a (1.48)	2.76 _b (1.26)
Feeling toward Interaction Group	56.97 _a (28.01)	59.00 _a (23.36)	62.74 _{ab} (25.11)	60.25 _a (19.17)	75.65 _b (18.98)
Identity-safety	3.85 _a (1.44)	4.01 _a (1.22)	4.12 _{ab} (1.20)	3.93 _a (1.18)	4.85 _b (1.19)

Note: Numbers in parentheses are cell standard deviations. Means not sharing a subscript within each dependent variable differ significantly at $p < .05$ after applying Bonferroni corrections.

the biased remark was directed, rather than an ingroup bystander. To address this confound, all confronters in Study 2 were bystanders of the biased remark.

As in Study 1, we hypothesized that, compared to participants in conditions without a biased comment, participants exposed to the biased comment would report more perpetrator offensiveness, less warmth toward the interaction group, and less identity-safety. We again tested competing predictions that a target group ally (i.e., a woman confronter) versus a dominant group ally (i.e., a man confronter) would be more effective at raising identity-safety.

Study 2 also introduced a manipulation of whether bystanders affirmed the confrontation. In line with group norms research (e.g., Reno et al., 1993), we reasoned that a single confronter may be unable to establish an egalitarian group norm, thereby providing an insufficient safety cue. A witness may wonder, do the other group members agree with the biased statement, or do they share the same egalitarian values as the confronter? Without more information, the target-group member may not be able to infer consensus about the confrontation's value within the group. Thus, we expected participants to report greater identity-safety when bystanders affirmed the confrontation compared to when they did not or when the comment was not confronted at all.

3.1 | Method

3.1.1 | Participants and design

Given the null effects for confrontation relative to No Confrontation in Study 1, we aimed to power Study 2 better through an increased number of participants and procedural improvements. G*Power (Faul et al., 2007) indicated that 426 participants (71 participants per cell) would provide 80% to detect a small-to-medium effect ($f = .18$).

Due to a programming error, initial data collection inadvertently left out the No Comment condition and replaced it with the No Confrontation condition (thus doubling up on data collection in this condition). Upon realizing the error toward the end of data collection, we corrected it to include data collection for the No Comment condition. For this reason, the cell size for the No Confrontation condition is higher ($n = 121$) than in the No Comment condition ($n = 72$).

Participants were 455 undergraduate women recruited via email to complete an online survey. As compensation, participants were entered into a lottery to win an Amazon gift card (one of 18 gift cards ranging from \$25 to \$100, with decreasing odds of winning as the amount increased). Based on a priori criteria, data were removed from five participants who did not give post-session consent to use their data; five participants who responded incorrectly to both manipulation-based attention checks (i.e., incorrectly reporting the condition they were in) and failed an audio test; and four participants who responded incorrectly to both manipulation-based attention checks and responded repetitively (e.g., all "4"s). Thus, 441 participants were included in the final dataset ($M_{\text{age}} = 20.06$, $SD_{\text{age}} = 1.96$; 72.34% White).

The between-participants design included six conditions. Four conditions included a sexist remark that was confronted by either a man or a woman, and this confrontation either was or was not affirmed by other members of the group. A fifth condition included a sexist remark that was not confronted. The final condition did not include a sexist remark.

3.1.2 | Procedure

The cover story and procedure were the same as in Study 1, except (a) participants listened to an audio file of the group interaction, (b) the biased comment and accompanying confrontation concerned gender instead of race, and (c) for some participants, the confrontation was affirmed by other members of the group.

The setting for the critical interaction was described as a workplace breakroom, with four coworkers interacting. The characters' names and voices conveyed gender. Participants were randomly assigned to one of six experimental conditions. In conditions including a biased comment, a White man stated, "I just don't think women are assertive enough to lead the IT team. And do you think a woman could actually stand up to the other directors? We don't have any female directors right now, and when we have had women they didn't last long". In conditions with a confrontation, a man or woman then said, "C'mon, women can be assertive and very strong leaders! Your statements sound a little unfair and sexist, don't you think?" In the two Confrontation Affirmed conditions, two remaining group members (one man and one woman) responded, "I agree. Let's not assume that women make poor directors", and "Yeah, hiring for any type of job should have nothing to do with gender". In the No Confrontation condition, the biased remark was left unconfronted. In the No Comment condition, no biased remark was made. At the end, the audio faded out, so that participants did not know what else was said.

3.2 | Results

3.2.1 | Analytic approach

Initially, feeling and identity-safety were predicted in single-factor (No Confrontation, Female Confrontation without Affirmation, Male Confrontation without Affirmation, Affirmed Female Confrontation, Affirmed Male Confrontation, and No Comment) between-participants ANCOVAs. These analyses revealed a significant main effect for condition for each dependent variable ($F_s > 5.88$, $p_s < .001$). However, follow-up comparisons with Bonferroni corrections did not reveal any instance in which the male and female confronter conditions differed, regardless of affirmation condition ($p_s = 1.00$). Thus, as in Study 1, confronter identity played no role in the results.

Given these non-significant effects, we collapsed across confronter gender in subsequent analyses. We analyzed each dependent variable with a single-factor between-participants variable with four levels: No Confrontation, Confrontation without Affirmation,

TABLE 4 Descriptive statistics, reliability, and correlations between measures, Study 2

	<i>M (SD)</i>	α	1.	2.
1. Perpetrator Offensiveness	5.08 (1.44)	.90	-	
2. Feeling toward Interaction Group	64.47 (22.21)	-	-.26*	-
3. Identity-safety	4.58 (1.12)	.96	-.28*	.83*

Note: Note: Feeling toward the interaction group was a one-item measure, so no alpha is reported.

* $p \leq .001$.

Confrontation Affirmed, and No Comment (i.e., sexist remark was not made).⁴ As in Study 1, perpetrator offensiveness was predicted with an ANOVA; feeling and identity-safety were predicted with an ANCOVA.

See Table 4 for descriptive statistics, reliability, and inter-measure correlations, and Table 5 for cell means and standard deviations.

3.2.2 | Perpetrator offensiveness

The effect of condition was significant, $F(3, 437) = 67.93, p < .001, \eta_p^2 = .32$. As with Study 1, the biased comment was interpreted as intended. The perpetrator was seen as more offensive in conditions with a biased comment compared to the No Comment condition, regardless of whether the comment was not confronted, $t(437) = 12.83, SE = 0.18, p < .001, 95\%CI [-2.75, -1.81]$, confronted, $t(437) = 11.95, SE = 0.18, p < .001, 95\%CI [-2.59, -1.65]$, or confronted and affirmed, $t(437) = 12.30, SE = 0.18, p < .001, 95\%CI [-2.63, -1.70]$. No other comparisons were significant, $p_s = 1.00$.

3.2.3 | Feeling toward interaction group

Feeling toward the filler interaction group was not a significant covariate, $F(1, 436) = 0.67, p = .41, \eta_p^2 = 0.002$. More importantly, the main effect for condition was significant, $F(3, 436) = 11.36, p < .001, \eta_p^2 = 0.07$. As in Study 1, participants felt warmer toward the interaction group when no biased comment was made than when a biased comment was made but not confronted, $t(436) = 5.73, SE = 3.20, p < .001, 95\%CI [-26.78, -9.84]$. Furthermore, the Confrontation condition again did not differ from the No Confrontation condition, $t(436) = 1.60, SE = 2.76, p = .66, 95\%CI [-11.72, 2.89]$, and was significantly less warm than the No Comment condition, $t(436) = 4.35,$

$SE = 3.19, p < .001, 95\%CI [-22.36, -5.44]$. When the confrontation was affirmed by others in the group, feelings were warmer than in the No Confrontation condition, although this difference did not reach the Bonferroni-corrected significance level, $t(436) = 2.62, SE = 2.74, p = .054, 95\%CI [-14.43, 0.07]$. Nevertheless, feelings toward the interaction group were still colder with an affirmed confrontation than in the No Comment condition, $t(436) = 3.51, SE = 3.17, p = .003, 95\%CI [-19.54, -2.72]$.

3.3 | Identity-safety

Filler identity-safety was not a significant covariate, $F(1, 436) = 0.01, p = .91, \eta_p^2 < .001$. As expected, the main effect for condition was significant, $F(3, 436) = 15.68, p < .001, \eta_p^2 = .10$. Participants felt more identity-safety in the No Comment condition than in the No Confrontation condition, $t(436) = 6.58, SE = 0.16, p < .001, 95\%CI [-1.47, -0.62]$. We again saw little support for the idea that confrontation alone could improve identity-safety. Identity-safety in the Confrontation condition did not differ from the No Confrontation condition, $t(436) = 2.27, SE = 0.14, p = .14, 95\%CI [-0.05, 0.68]$, and was significantly lower than in the No Comment condition, $t(436) = 4.63, SE = 0.16, p < .001, 95\%CI [-1.16, -0.32]$. When the confrontation was affirmed, greater identity-safety was reported compared to no confrontation, $t(436) = 4.16, SE = 0.14, p < .001, 95\%CI [0.21, 0.93]$. Nevertheless, identity-safety was still significantly lower in the Confrontation Affirmed condition than in the No Comment condition, $t(436) = 3.04, SE = 0.16, p = .02, 95\%CI [-0.90, -0.06]$. These findings suggest that affirmed confrontations signal identity-safety in the group, although not to the same degree as if no biased comment was made.

3.4 | Summary

A sexist comment among a group of coworkers caused women to experience less identity-safety. As in Study 1, when the remark was simply confronted (whether by a man or a woman), identity-safety was not restored. However, when other group members affirmed a confrontation by likewise indicating that the sexist comment was unacceptable, identity-safety was significantly boosted compared to no confrontation. Thus, Study 2 suggests that, for a safety cue to be established following an act of sexism, there must be group consensus for the confrontation, which presumably makes clear that social norms within the group are against the expression of sexism.

4 | STUDY 3

Study 3 examined whether the effect of affirmed confrontations generalizes to anti-Black racism. We considered two possible outcomes. First, as in Study 2, an affirmed confrontation will restore Black individuals' identity-safety to the same extent that affirmed confrontation restored identity-safety for women. However,

⁴We also repeated all analyses with more stringent exclusion criteria, excluding 29 participants who (a) incorrectly responded to two out of three number-based attention checks, (b) incorrectly responded to one out of two of the manipulation-based attention checks, or (c) did not pass the audio test. Excluding these participants affected only one result: the simple effect between "No Comment" and "Affirmed Confrontation" for desire to join became significant at $p = .03$. Given this minor difference, we report the main analyses including these participants.

TABLE 5 Measures as a function of confrontation condition, Study 2

Measure	No Confrontation <i>n</i> = 121	Confrontation (no affirmation) <i>n</i> = 122	Confrontation affirmed <i>n</i> = 126	No comment <i>n</i> = 72
Perpetrator Offensiveness	5.52 _a (1.40)	5.36 _a (1.14)	5.41 _a (1.11)	3.25 _b (1.04)
Feeling toward Interaction Group	58.17 _a (22.27)	62.64 _{ab} (21.86)	65.44 _{ab} (22.03)	76.50 _c (18.13)
Identity-safety	4.16 _a (1.14)	4.47 _{ab} (1.00)	4.72 _b (1.15)	5.21 _c (0.89)

Note: Numbers in parentheses are cell standard deviations. Means not sharing a subscript within each dependent variable differ significantly at $p < .05$ after applying Bonferroni corrections. Because we collapsed across male and female confronters to create a single “Confrontation (no affirmation) condition” and “Confrontation Affirmed” condition, these conditions have slightly higher *n*s than the “No Comment” condition. Furthermore, the No Confrontation condition has a higher *n* than the No Comment condition due a computer programming error.

societal norms against racism toward Blacks are stronger than societal norms against sexism (Fiske & Stevens, 1993; Woodzicka, Mallett, Hendricks, & Pruitt, 2015). Given this, we also considered a second possibility that the expression of a biased statement about Black people may be perceived as especially egregious, so that, regardless of whether a confrontation is affirmed, it will not restore identity valuation within the group. This competing hypothesis thus posits that the damage done by an anti-Black statement cannot be countered through confrontation, even if it is affirmed by other group members.

Our two competing hypotheses for Study 3, along with methods and data analysis plans, were preregistered, <https://osf.io/cmn4v>.

4.1 | Method

G*Power v.3.1.9.2 (Faul et al., 2007) indicated that 467 participants (or, approximately 58 participants per cell) would provide 90% power to detect a small-to-medium sized effect ($f = .18$). We tested our hypotheses among White and Black women. Given their underrepresentation, Black women are difficult to recruit, so we optimized statistical power by over-sampling White women (Hennes, Lane, & Neo, 2019).

Black and White women ($N = 461$) received \$1.00 for participating through TurkPrime. Data from six participants were removed: two who did not give post-session consent to analyze their data, two who experienced technical issues, and two who missed both manipulation-based attention checks and did not pass the audio test. The final dataset included 455 participants ($M_{age} = 41.26$, $SD_{age} = 12.67$; 59.6% White, 40.4% Black).

Study 3 used a 2 (Bias Type: Race vs. Gender) \times 4 (Condition: No Confrontation, Confrontation Only, Affirmed Confrontation, and No Biased Comment) design. We recruited both Black and White women, assigning Black participants to the race condition, White participants to the gender condition, and randomly assigning confrontation condition. White participants were older (20–73; $M = 43.33$, $SD = 12.93$) than Black participants (18–67, $M = 38.21$, $SD = 11.67$), $t(453) = 4.43$, $p < .001$, and Black participants were more liberal ($M = 2.98$, $SD = 1.49$) than White participants ($M = 3.69$, $SD = 1.81$) on a 1 (Very Liberal) to 7 (Very Conservative) scale, $t(453) = 4.10$, $p < .001$. However, including age and political orientation as covariates did not change results. Also, White and Black participants did not differ for any outcome

variables in the No Comment condition ($ps > .63$), indicating that they responded similarly with a neutral interaction.

4.1.1 | Procedure

The procedure and measures were identical to Study 2, with a few exceptions. In the critical audio, we manipulated whether the biased comment concerned women or Black people and used the following wording: “Do you really think that a [woman/Black person] could be up to it? It is central to the company's functioning, and requires a lot of advanced qualifications...” Given the non-significant confronter identity effects in Studies 1 and 2, we did not manipulate this variable. A White man always confronted the biased comment. Finally, we did not collect measures of perceptions of the confronter.⁵

4.2 | Results

4.2.1 | Analytic approach

Each dependent variable was predicted using Type of Bias (Race vs. Gender) and Confrontation Condition (No Confrontation, Confrontation Only, Affirmed Confrontation, and No Biased Comment) in a between-participants ANCOVA.⁶ See Table 6 for de-

⁵As an exploratory measure, participants also reported their stigma consciousness of both racism and sexism, so that these items could be included as covariates in the ANOVA models. Although there were significant effects of gender stigma consciousness, but not racism stigma consciousness, on feeling and identity-safety (see Appendix S1), these variables did not change the main effect of confrontation condition, nor the interaction between confrontation condition and bias type, for either variable. Participants also reported the extent to which they believe sexism and racism are acceptable in today's society, which we explored as a potential mechanism between confrontation condition and identity-safety. The effect of confrontation condition and the interaction between confrontation condition and bias type were not significant (see Appendix S1). Given this, we did not conduct subsequent mediation analyses. See Appendix S1 for a more detailed consideration of why this potential mechanism was not significant.

⁶We conducted additional analyses excluding participants ($n = 49$) who (a) did not pass the attention checks, (b) showed evidence of repetitive responding (e.g., responding “7” to every question), or (c) suspected the study hypotheses. Excluding these participants did not change results, so, based on an a priori, pre-registered decision, we report the main analyses including these participants. We also included three checks to ensure “server farms” (TurkPrime, 2018) were not completing the study, which all participants passed.

TABLE 6 Descriptive statistics, reliability, and correlations between measures, Study 3

	M (SD)	α	1.
1. Feeling toward Interaction Group	63.61 (24.99)	-	-
2. Identity-safety	4.40 (1.44)	.98	.84*

Note: Note: Feeling toward the interaction group was a one-item measure, so no alpha is reported.

* $p \leq .001$.

scriptive statistics, reliability, and correlations, and Table 7 for condition cell means and standard deviations.

4.2.2 | Feeling toward the interaction group

Feeling toward the filler interaction group was a significant covariate, $F(1, 446) = 21.33, p < .001, \eta_p^2 = .12$. Furthermore, the main effect for bias type was significant, $F(1, 446) = 6.86, p = .009, \eta_p^2 = .02$. Participants in the Race Bias condition ($M = 60.30, SD = 27.48$) felt colder toward the interaction group than participants in the Gender Bias condition ($M = 66.14, SD = 22.63$).

We also observed the expected main effect of confrontation condition, $F(3, 446) = 19.90, p < .001, \eta_p^2 = .12$, which was not moderated by an interaction between type of bias and confrontation condition, $F(3, 446) = 1.73, p = .16, \eta_p^2 = .01$. The results mirrored Study 2. Participants reported warmer feelings with no comment than no confrontation, $t(446) = 7.42, SE = 3.10, p < .001, 95\%CI [14.80, 31.22]$. Furthermore, the Confrontation condition did not differ from the No Confrontation condition, $t(446) = 1.94, SE = 3.07, p = .32, 95\%CI [-14.09, 2.17]$, and was significantly colder than the No Comment condition, $t(446) = 5.56, SE = 3.07, p < .001, 95\%CI [8.93, 25.17]$. However, when the confrontation was affirmed, feelings towards the interaction group improved compared to the No Confrontation condition, $t(446) = 2.80, SE = 3.13, p = .03, 95\%CI [-17.08, -0.48]$.⁷ Still, feelings in the Affirmed Confrontation condition were significantly colder than in the No Comment condition, $t(446) = 4.56, SE = 3.12, p < .001, 95\%CI [5.97, 22.49]$.

4.2.3 | Identity-safety

Filler identity-safety was a significant covariate, $F(1, 446) = 14.10, p < .001, \eta_p^2 = .03$. Furthermore, the main effect for bias type was significant, $F(1, 446) = 11.20, p = .001, \eta_p^2 = .02$. Participants in the Race Bias condition ($M = 4.14, SD = 1.44$) reported less identity-safety than participants in the Gender Bias condition ($M = 4.59, SD = 1.37$).

⁷When the covariate was excluded, this comparison was not significant, $t(446) = 2.54, SE = 3.20, p = .07, 95\%CI [-0.34, 16.62]$, though it still trended in the hypothesized direction.

More importantly, we observed the expected main effect of confrontation condition, $F(3, 446) = 17.49, p < .001, \eta_p^2 = .11$, which was not moderated by an interaction between type of bias and confrontation condition, $F(3, 446) = 1.31, p = .27, \eta_p^2 = .009$. Replicating Studies 1 and 2, participants felt more identity-safety in the No Comment condition than in the No Confrontation condition, $t(446) = 7.07, SE = 0.18, p < .001, 95\%CI [0.79, 1.74]$. Furthermore, the Confrontation Only condition did not differ from the No Confrontation condition, $t(446) = 2.22, SE = 0.18, p = .16, 95\%CI [-0.87, 0.08]$, and was significantly lower than the No Comment condition, $t(446) = 4.91, SE = 0.18, p < .001, 95\%CI [0.40, 1.34]$. Replicating Study 2, affirmed confrontation boosted identity-safety compared to no confrontation, $t(446) = 3.17, SE = 0.18, p = .01, 95\%CI [-1.05, -0.09]$, but identity-safety was still higher in the No Comment than the Affirmed Confrontation condition, $t(446) = 3.84, SE = 0.18, p = .001, 95\%CI [0.21, 1.17]$.

4.3 | Summary

A group-affirmed confrontation helped to restore participants' identity-safety to the same extent for Black women who were exposed to a racist comment as for White women who were exposed to a sexist comment. Thus, even in the case of racism toward Black people, target group members will be able to experience warmer feelings and greater identity-safety if biased comments are confronted and the confrontation is affirmed by others.

5 | INTEGRATIVE DATA ANALYSIS

We examined our hypotheses across all studies using integrative data analyses (IDA), which has been called the "gold standard meta-analytic approach" (McShane & Böckenholt, 2017). We combined the data from all studies and conducted separate one-way ANCOVAs testing the effect of (a) confronter identity (ally vs. in-group confrontation) and (b) confrontation (No Confrontation, Confrontation Only, Confrontation Affirmed, and No Comment) on feeling toward the interaction group and identity-safety. In accordance with IDA best practices (Curran & Hussong, 2009), we statistically accounted for which study the data came from by including study as a covariate. The relevant variable from the filler script was again entered as a covariate ($ps < .002$), and pairwise comparisons were Bonferroni-corrected.

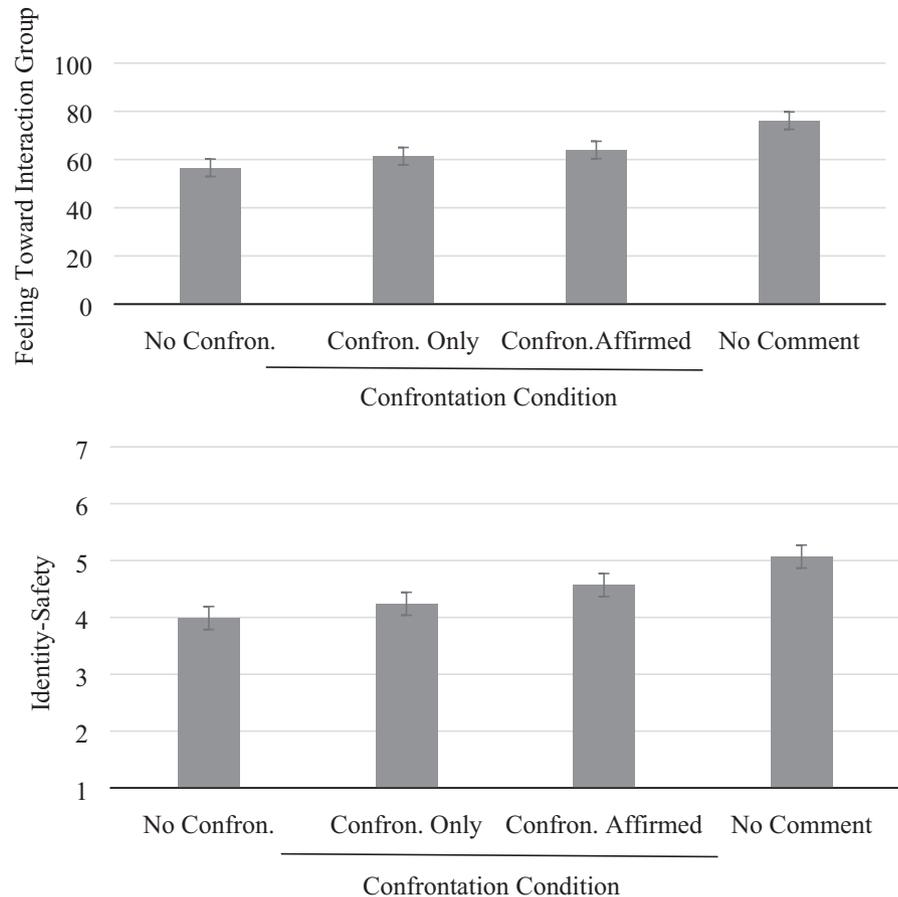
5.1 | Feeling toward the interaction group

In the analysis examining confronter identity, feelings toward the filler interaction groups were a significant covariate when predicting feelings toward the critical interaction groups, $F(1, 611) = 9.45, p = .002, \eta_p^2 = .02$. However, the main effect of ally condition was not significant, $F(1, 611) = 0.45, p = .51, \eta_p^2 = .001$. Thus, there was

TABLE 7 Measures as a function of confrontation condition, Study 3

Measure	No Confrontation <i>n</i> = 112	Confrontation (no affirmation) <i>n</i> = 119	Confrontation affirmed <i>n</i> = 108	No comment <i>n</i> = 116
Feeling toward Interaction Group	54.72 _a (25.17)	61.03 _{ab} (24.73)	62.28 _b (23.38)	76.09 _c (21.78)
Identity-safety	3.85 _a (1.36)	4.28 _{ab} (1.42)	4.39 _b (1.40)	5.06 _c (1.31)

Note: Note: Numbers in parentheses are cell standard deviations. Means not sharing a subscript within each dependent variable differ significantly at $p < .05$ after Bonferroni corrections were applied.

FIGURE 2 Measures as a function of confrontation condition, Integrative Data Analysis. Note. Error whiskers represent standard errors

no evidence that participants' feelings depended on confronter identity.

In the analysis examining confrontation condition, feelings toward the filler interaction groups were again significant, $F(1, 1,103) = 22.98, p < .001, \eta_p^2 = .02$. More importantly, the main effect for confrontation was significant, $F(3, 1,103) = 34.39, p < .001, \eta_p^2 = .09$ (see Figure 2). Confrontation alone did not boost warmth toward the interaction group compared to the No Confrontation condition, $t(1,103) = 2.19, SE = 1.83, p = .17, 95\%CI [-0.84, 8.84]$, and reported feelings were significantly less warm in the Confrontation Only than in the No Comment condition, $t(1,103) = 8.10, SE = 1.92, p < .001, 95\%CI [-20.63, -10.48]$. However, as in Studies 2 and 3, a group-affirmed confrontation boosted warmth compared to no confrontation, $t(1,103) = 5.62, SE = 2.10, p < .001, 95\%CI [-17.31, -6.24]$, but feelings were still colder than when no biased comment was made, $t(1,103) = 3.85, SE = 2.02, p = .001, 95\%CI [2.45, 13.13]$.

5.2 | Identity-safety

In the analysis examining confronter identity, filler identity-safety was not a significant covariate, $F(1, 611) = 2.40, p = .12, \eta_p^2 = .004$. The main effect of ally condition was not significant, $F(1, 611) = 2.12, p = .15, \eta_p^2 = .003$. Thus, there was again no evidence that identity-safety depended on confronter identity.

In the analysis involving confrontation condition, filler identity-safety was a significant covariate, $F(3, 1,104) = 10.95, p = .001, \eta_p^2 = .01$. The main effect for confrontation was significant, $F(3, 1,104) = 35.40, p < .001, \eta_p^2 = .09$ (see Figure 2). Again, confrontation alone did not increase identity-safety compared to no confrontation, $t(1,104) = 2.44, SE = 0.10, p = .09, 95\%CI [-0.02, 0.51]$, and identity-safety was significantly lower in the Confrontation condition than in the No Comment condition, $t(1,104) = 7.90, SE = 0.11,$

$p < .001$, 95%CI [-1.11, -0.56]. More importantly, a group-affirmed confrontation increased identity-safety compared to no confrontation, $t(1,104) = 5.33$, $SE = 0.11$, $p < .001$, 95%CI [0.29, 0.88], and to confrontation only, $t(1,104) = 3.23$, $SE = 0.11$, $p = .008$, 95%CI [0.06, 0.62]. However, identity-safety was still lower in the Affirmed Confrontation condition compared to the No Comment condition, $t(1,104) = 4.28$, $SE = 0.12$, $p < .001$, 95%CI [-0.80, -0.19].

In sum, the integrative data analysis provided further evidence that affirmed confrontation can promote warmth and identity-safety after a biased comment, though still lower than if a biased comment had been made at all.

6 | GENERAL DISCUSSION

Marginalized group members often witness biased remarks being made about their group (e.g., Swim et al., 2003), which can threaten their identity and decrease feelings of belonging and safety within the given environment (e.g., Murphy et al., 2007). The present research investigated whether ally confrontations can restore target group members' safety and belonging following such biased comments. Across three studies, the present research demonstrates that, regardless of confronter identity (i.e., ally or ingroup member), confrontations caused target group participants to feel like their identity was safe and valued following a biased remark, but only if other members of the interaction group spoke up to affirm the confrontation.

In Studies 1 and 2, we tested whether confronting a biased comment would increase positive feelings toward the interaction group and identity-safety compared to when no confrontation occurred. Furthermore, we tested competing hypotheses concerning whether dominant group ally or ingroup confrontations would provide the most effective safety cue. However, results from both studies demonstrated that target group members saw little value in confrontations performed by lone confronters (i.e., confrontations that were not affirmed by others). Regardless of whether the confrontation was performed by someone having the same social identity as the person targeted by bias (i.e., an ingroup member), by someone with a different but still marginalized identity (i.e., a marginalized group ally), or by someone with a dominant group identity (i.e., a dominant group ally), these lone confrontations did not restore identity-safety following a biased remark. Studies 2 and 3, which compared lone confrontations to affirmed confrontations, demonstrated that bystander social support is necessary for confrontations to serve as a safety cue. Specifically, confrontation boosted identity-safety following a biased remark if it was supported by other members of the interaction group via verbal affirmations that agreed with the confrontation content. Thus, overall, the present research suggests that affirmed confrontations serve as a safety cue for target group individuals following a biased comment, regardless of confronter identity. These conclusions were supported by an integrative data analysis that combined the data from all three studies.

These findings contribute to the existing literature on target group members' reactions to allies. Previous research has found that target group members sometimes regard allies less positively than ingroup members who stand up against bias. For example, target group members perceive allies as less aware of inequality (Brown & Ostrove, 2013), less supportive of target group goals, less capable of supportive contact (i.e., supporting the group while upholding their autonomy; Droogendyk et al., 2016; Wiley & Dunne, 2019), and as having less solidarity (Craig & Richeson, 2012; Warner et al., 2014) compared to ingroup members. On the other hand, other research on the interpersonal consequences of ally confrontation has found that allies are well-liked by target group members and are perceived as responding to the situation appropriately (Kutlaca et al., 2020). The present research contributes to this existing literature by demonstrating that confrontations by allies that were affirmed by bystanders promoted feelings of warmth and identity-safety to the same extent as ingroup confrontations. Thus, although allies may not always be viewed as positively as ingroup members, their confrontation of bias can signal safety to target group members in interpersonal situations (see also Cihangir et al., 2014).

The present research also adds to the literature on safety cues. Existing safety cue work has examined the extent to which both environmental features (such as stereotypically masculine objects; Chaney & Sanchez, 2018a, 2018b; Cheryan et al., 2009) and the presence of other people (such as role models, or non-White friends; Pietri, Johnson, & Ozgumus, 2018; Wout et al., 2014) contributes to threat or safety within a given environment. Yet, little research has studied factors that influence the efficacy of those safety cues, nor have these issues been examined in the context of confrontation. The present research highlights confrontations as a potential safety cue and affirmations as a factor that influences effectiveness.

The present work also extends prior confrontation research focusing on bias reduction among dominant group members (e.g., Chaney & Sanchez, 2018a; Czopp et al., 2006). In contrast to our findings showing the necessary role of bystanders who affirm a confrontation, single confronters have generally succeeded in reducing dominant group members' expressions of bias (for a review, see Czopp, 2019). This discrepancy may hinge on different psychological perspectives and concerns across the two situations. People who are confronted about expressing bias may perceive minimal cost to complying. Furthermore, reducing subsequent expressions of bias may bring benefits, such as fulfilling affiliative and epistemic needs by participating in a non-biased, "shared reality" (Echterhoff, Higgins, & Levine, 2009). In contrast, from the perspective of the target group members in our research, there likely was little perceived benefit to joining an interaction group in which a biased comment had been made. Furthermore, the threat of further bias by the perpetrator and possibly others in the interaction group may be perceived as significant potential costs, unless unified opposition to bias is expressed through confrontation affirmations. Thus, the differential balance of benefits and costs across the two situations may explain why one confronter is

enough for prejudice reduction and not enough for the establishment of safety cues.

This balance of benefits and costs may also explain why affirmed confrontations could not completely restore feeling and identity-safety following a biased remark. Even though feeling and identity-safety increased following an affirmed confrontation compared to when the biased comment was not confronted, both outcomes still remained significantly lower than if no biased comment had been made at all. Likely, the target group member still perceived a significant amount of threat from the perpetrator, even though the other interaction group members expressed opposition to the biased comment. Perhaps if the interaction group had more extremely censured the perpetrator, such as ostracizing them from the group, feeling and identity-safety would have been completely restored compared to the No Comment condition. However, such extreme action was not taken in the present paradigm. In the present paradigm, the cost of joining a group with a biased member likely outweighed any potential benefit. As a result, participants reported less positive feeling and identity-safety following a biased comment, even when the comment was confronted and affirmed by others in the interaction group.

Even though affirmed confrontations could not restore feeling and identity-safety completely, they still boosted identity-safety compared to when the comment was not confronted or when the comment was confronted by one person. Why was affirmed confrontation a better safety cue than lone confrontation? When the other interaction group members do not respond to the confrontation, target group members may question the extent to which the confrontation represents group norms. However, when additional interaction group members affirm the confrontation, the impact of the confrontation increases; target group members may be more likely to believe that the confrontation represents egalitarian group norms. The effect of affirmed versus not affirmed confrontation aligns with the literature on group norms. For instance, Social Impact Theory (Latané, 1981) states that as the number of people supporting a norm increases (such as through affirmation), the social impact of the norm also increases. In other words, if multiple people speak up supporting a condemnation of prejudice, then both a descriptive (i.e., what people do) and injunctive (i.e., what people should do) egalitarian norm is established (e.g., Cialdini et al., 1990; Reno et al., 1993). As a result, target group members should feel safer in the given environment and believe that their identity will be supported and valued within the given environmental context following an affirmed confrontation (Ata et al., 2009; Miller et al., 2000; Smith & Louis, 2008). Although the group norms literature supports this line of thinking, we did not test this process directly. Future research should explore the possible mediating role of group norms and social support.

Although the present research highlights the benefits of affirmed confrontation, people may be hesitant to speak up following another person's confrontation. According to the Confronting Prejudiced Responses (CPR) model (Ashburn-Nardo, Morris, & Goodwin, 2008), bystanders are much more likely to confront prejudice when they feel a sense of responsibility to confront, and they may feel less

responsibility when someone else has already confronted. Therefore, allies should be encouraged both to confront and to affirm others' confrontations, because doing so may establish strong egalitarian norms and afford greater identity-safety for target group members.

6.1 | Limitations and future research

The present research examined confrontations as a safety-cue across several different forms of prejudice (i.e., anti-Asian, gender, and anti-Black bias) with marginalized-group participants (i.e., Chinese-American, White women, and Black women) who varied in age (i.e., older participants recruited through Qualtrics and TurkPrime, younger undergraduate college students). Despite this diversity, the present research is limited in that it examines each marginalized group individually. Many people hold a combination of intersecting identities, and our methods did not examine potential intersectional effects. For example, in Study 3, Black women only responded to confrontations of anti-Black bias and not to confrontations of sexism. This decision was made strategically, as we anticipated that we would be unable to recruit enough Black women participants to examine their reactions to both sexism and racism. Yet, this decision also discounts the unique experience of Black women as both gender and racial minorities. Black women do in fact experience racism and sexism simultaneously (Purdie-Vaughns & Eibach, 2008; Remedios & Snyder, 2015), and, like other women of color, subsequently face unique forms of discrimination due to their intersecting marginalized identities (e.g., Major & Crocker, 1993; Mohr & Purdie-Vaughns, 2015; Purdie-Vaughns & Eibach, 2008). For example, Black women are subject to both intersectional invisibility, in which they are ignored and underrepresented compared to prototypical, singly stigmatized group members (Purdie-Vaughns & Eibach, 2008), and double jeopardy, in which they experience more discrimination than prototypical members of their groups (Beal, 2008; Remedios & Snyder, 2015). Perhaps this is why Black women reported lower identity-safety (across all conditions) than White women (Study 3). Given these issues, it is critical that future research examine affirmed confrontations as a safety cue with an intersectional approach.

The paradigms we used may have limitations that should also be examined in future research. For example, people may respond differently in an actual group setting, where discussion is face-to-face rather than reading about (Study 1) or listening to (Studies 2 and 3) a group interaction. In line with Social Impact Theory (e.g., Latané & Wolf, 1981), such immediate, individualized interaction may increase the impact of both the confrontation and the subsequent interaction group affirmations. Furthermore, future research would benefit from incorporating behavioral measures, such as nonverbal behavior.

7 | CONCLUSION

At the beginning of this article, we asked whether Andy Murray's confrontation provided a safety cue for Black women in the Olympic

environment. Importantly, his confrontation prompted an outpouring of support, including Tweets such as “I award [Murray] the gold medal for feminism in the men's events” (Phillips, 2016) and “[Murray] just reminded [the interviewer] that women are people, too” (Rowling, 2016). While Murray's confrontation alone may have not been enough to act as a safety cue, the present research suggests that, when combined with the deluge of online affirmations, identity-safety for Black women within the Olympic environment may have been improved. This serves as just one example of how allies may provide benefits through their confrontations, especially when others speak up to affirm the ally's actions. Specifically, when confrontations of prejudice are affirmed, they may provide an identity-safety cue for members of groups being targeted by bias, increasing their belonging and safety in an environment. We hope that the present research findings spur more work on the role of confrontation and allyship in boosting identity-safety among marginalized individuals.

ACKNOWLEDGEMENTS

The first author's efforts on this work were supported by a Graduate Research Fellowship from the National Science Foundation (DGE-1333468). Study 2 and Study 3 were supported by a grant from the Purdue University Acceptance and Inclusion Consortium awarded to the second author. We thank Anthony Kowalkowski for help with various aspects of this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

TRANSPARENCY STATEMENT

All study materials, including the scripts, syntaxes, outputs, Study 3's pre-registration, and results of additional and exploratory analyses (i.e., Appendix S1), are available as Supplemental Materials and can be found on the Open Science Framework's website (see <https://osf.io/jqf2c/>). Raw data is available upon request from the corresponding author; it cannot be made publicly available since participants were informed that data would be accessible only to those associated with the project.

ORCID

Laura K. Hildebrand  <https://orcid.org/0000-0002-3236-5624>

REFERENCES

- Affleck, G., & Tennen, H. (1996). Construing benefits from adversity: Adaptational significance and dispositional underpinnings. *Journal of Personality, 64*(4), 899–922. <https://doi.org/10.1111/j.1467-6494.1996.tb00948.x>
- Allport, G. W. (1954). *The nature of prejudice*. New York, NY: Addison-Wesley.
- Ashburn-Nardo, L., Morris, K. A., & Goodwin, S. A. (2008). The Confronting Prejudiced Responses (CPR) model: Applying CPR in organizations. *Academy of Management Learning and Education, 7*, 332–342. <https://doi.org/10.5465/AMLE.2008.34251671>
- Ata, A., Bastian, B., & Lusher, D. (2009). Intergroup contact in context: The mediating role of social norms and group-based perceptions on the contact–prejudice link. *International Journal of Intercultural Relations, 33*(6), 498–506. <https://doi.org/10.1016/j.ijintrel.2009.05.001>
- Barr, S. (2019). Andy Murray's ten best feminist moments. *The Independent*. Retrieved from <https://www.independent.co.uk/life-style/women/andy-murray-wimbledon-2019-feminist-moments-equal-pay-sexism-women-serena-williams-a8996971.html>
- Beal, F. M. (2008). Double jeopardy: To be black and female. *Meridians, 8*(2), 166–176. <https://doi.org/10.2979/MER.2008.8.2.166>
- Becker, J. C., & Barreto, M. (2019). Personal, collective, and group-distancing motives underlying confrontation of prejudice. In R. Mallet & M. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors* (pp. 141–158). London, UK: Academic Press.
- Branscombe, N. R., Schmitt, M. T., & Harvey, R. (1999). Perceiving pervasive discrimination among African Americans: Implications for group identification and well-being. *Journal of Personality and Social Psychology, 77*(1), 135–149. <https://doi.org/10.1037/0022-3514.77.1.135>
- Brown, K. T. (2015). Perceiving allies from the perspective of non-dominant group members: Comparisons to friends and activists. *Current Psychology, 34*(4), 713–722. <https://doi.org/10.1007/s12144-014-9284-8>
- Brown, K. T., & Ostrove, J. M. (2013). What does it mean to be an ally? The perception of allies from the perspective of people of color. *Journal of Applied Social Psychology, 43*(11), 2211–2222. <https://doi.org/10.1111/jasp.12172>
- Burns, M. D., & Monteith, M. J. (2019). Confronting stereotypic biases: Does internal versus external motivational framing matter? *Group Processes & Intergroup Relations, 22*(7), 930–946. <https://doi.org/10.1177/1368430218798041>
- Chaney, K. E., & Sanchez, D. T. (2018a). The endurance of interpersonal confrontations as a prejudice reduction strategy. *Personality and Social Psychology Bulletin, 44*(3), 418–429. <https://doi.org/10.1177/0146167217741344>
- Chaney, K. E., & Sanchez, D. T. (2018b). Gender-inclusive bathrooms signal fairness across identity dimensions. *Social Psychological and Personality Science, 9*(2), 245–253. <https://doi.org/10.1177/1948550617737601>
- Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045–1060. <https://doi.org/10.1037/a0016239>
- 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>
- Cihangir, S., Barreto, M., & Ellemers, N. (2014). Men as allies against sexism: The positive effects of a suggestion of sexism by male (vs. female) sources. *SAGE Open, 4*(2), 1–12. <https://doi.org/10.1177/2158244014539168>
- Craig, M. A., & Richeson, J. A. (2012). Coalition or derogation? How perceived discrimination influences intraminority intergroup relations. *Journal of Personality and Social Psychology, 102*(4), 759–777. <https://doi.org/10.1037/a0026481>
- Craig, M. A., & Richeson, J. A. (2016). Stigma-based solidarity: Understanding the psychological foundations of conflict and coalition among members of different stigmatized groups. *Current Directions in Psychological Science, 25*(1), 21–27. <https://doi.org/10.1177/0963721415611252>
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology, 82*(3), 359. <https://doi.org/10.1037/0022-3514.82.3.359>
- Crandall, C. S., & Stangor, C. (2005). Conformity and prejudice. In J. Dovidio, P. Glick & L. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 295–310). Malden, MA: Blackwell Publishing.

- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, 14(2), 81–100. <https://doi.org/10.1037/a0015914>
- Czopp, A. M. (2019). The consequences of confronting prejudice. In R. K. Mallet & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors* (pp. 201–221). London, UK: Academic Press.
- Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, 29(4), 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 conflict. *Journal of Personality and Social Psychology*, 90(5), 784–803. <https://doi.org/10.1037/0022-3514.90.5.784>
- Droogendyk, L., Wright, S. C., Lubensky, M., & Louis, W. R. (2016). Acting in solidarity: Cross-group contact between disadvantaged group members and advantaged group allies. *Journal of Social Issues*, 72(2), 315–334. <https://doi.org/10.1111/josi.12168>
- Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues*, 70(4), 637–652. <https://doi.org/10.1111/josi.12083>
- Echterhoff, G., Higgins, E. T., & Levine, J. M. (2009). Shared reality: Experiencing commonality with others' inner states about the world. *Perspectives on Psychological Science*, 4(5), 496–521. <https://doi.org/10.1111/j.1745-6924.2009.01161.x>
- Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations*, 15(4), 487–502. <https://doi.org/10.1177/1368430211432894>
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Fiske, S. T., & Stevens, L. E. (1993). What's so special about sex? Gender stereotyping and discrimination. In S. Oskamp & M. Costanzo (Eds.), *Claremont symposium on applied social psychology* (vol. 6, pp. 173–196). Oakland, CA: Sage Publications Inc.
- Gervais, S. J., & Hillard, A. L. (2014). Confronting sexism as persuasion: Effects of a confrontation's recipient, source, message, and context. *Journal of Social Issues*, 70(4), 653–667. <https://doi.org/10.1111/josi.12084>
- Goffman, E. (1963). *Stigma; notes on the management of spoiled identity*. Englewood Cliffs, NJ: Prentice-Hall.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, 102(4), 700–717. <https://doi.org/10.1037/a0026659>
- 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>
- Hennes, E. P., Lane, S. P., & Neo, W. S. (2019). *Optimizing statistical power in studies of underrepresented populations by oversampling well-represented group members*. Data blitz presented at the Weary Symposium on Diversity and Social Identity, Columbus, OH.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- 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. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, 27(2), 254–263. <https://doi.org/10.1177/0146167201272010>
- Kutlaca, M., Becker, J., & Radke, H. (2020). A hero for the outgroup, a black sheep for the ingroup: Societal perceptions of those who confront discrimination. *Journal of Experimental Social Psychology*, 88, 103832. <https://doi.org/10.1016/j.jesp.2019.103832>
- Latané, B. (1981). The psychology of social impact. *American Psychologist*, 36, 343–356. <https://doi.org/10.1037/0003-066X.36.4.343>
- Latané, B., & Wolf, S. (1981). The social impact of majorities and minorities. *Psychological Review*, 88(5), 438–453. <https://doi.org/10.1037/0033-295X.88.5.438>
- Major, B., & Crocker, J. (1993). Social stigma: The consequences of attributional ambiguity. In D. Mackie & D. Hamilton (Eds.), *Affect, cognition and stereotyping: Interactive processes in group perception* (pp. 345–370). San Diego, CA: Academic Press.
- Mallet, R. K., & Monteith, M. J. (Eds.) (2019). *Confronting prejudice and discrimination: The science of changing minds and behaviors*. London, UK: Academic Press.
- Mallet, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequence of confronting sexism. *Journal of Experimental Social Psychology*, 47(1), 215–220. <https://doi.org/10.1016/j.jesp.2010.10.001>
- McShane, B. B., & Böckenholt, U. (2017). Single-paper meta-analysis: Benefits for study summary, theory testing, and replicability. *Journal of Consumer Research*, 43(6), 1048–1063. <https://doi.org/10.1093/jcr/ucw085>
- Miller, D. T., Monin, B., & Prentice, D. A. (2000). Pluralistic ignorance and inconsistency between private attitudes and public behaviors. In D. J. Terry & M. A. Hogg (Eds.), *Attitudes, behavior, and social context* (pp. 95–113). Mahwah, NJ: Erlbaum.
- Mohr, R. I., & Purdie-Vaughns, V. (2015). Diversity within women of color: Why experiences change felt stigma. *Sex Roles*, 73(9), 391–398. <https://doi.org/10.1007/s11199-015-0511-z>
- Murphy, M. C., Steele, C. M., & Gross, J. T. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, 18(10), 879–885. <https://doi.org/10.1111/j.1467-9280.2007.01995.x>
- Muthén, L., & Muthén, B. (2017). *Mplus (Version 8)[computer software]*. Los Angeles, CA: Muthén & Muthén.
- Nosek, B. A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science*, 16, 65–69. <https://doi.org/10.1111/j.1467-8721.2007.00477.x>
- Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology*, 74, 8–23. <https://doi.org/10.1016/j.jesp.2017.07.009>
- Phillips, J. (2016). I like Andy Murray even more for pointing out that women are people too. I award him the gold medal for feminism in the men's events. Twitter. Retrieved from https://twitter.com/jessphillips/status/765119275820613632?ref_src=twsrc%5Etfw
- Pietri, E. S., Johnson, I. R., & Ozigumus, E. (2018). One size may not fit all: Exploring how the intersection of race and gender and stigma consciousness predict effective identity-safe cues for Black women. *Journal of Experimental Social Psychology*, 74, 291–306. <https://doi.org/10.1016/j.jesp.2017.06.021>
- Purdie-Vaughns, V., & Eibach, R. (2008). Intersectional invisibility: The distinctive advantages and disadvantages of multiple subordinate-group identities. *Sex Roles*, 59(5), 377–391. <https://doi.org/10.1007/s11199-008-9424-4>
- Purdie-Vaughns, V., Steele, C. M., Davies, P. G., Dittmann, R., & Crosby, J. R. (2008). Social identity contingencies: How diversity cues signal threat or safety for African Americans in mainstream institutions.

- Journal of Personality and Social Psychology*, 94(4), 615. <https://doi.org/10.1037/0022-3514.94.4.615>
- Rasinski, H. M., & Czopp, A. M. (2010). The effect of target status on witnesses' reactions to confrontation of bias. *Basic and Applied Social Psychology*, 32(1), 8–16. <https://doi.org/10.1080/01973530903539754>
- Remedios, J., & Snyder, S. (2015). Where do we go from here? Toward an inclusive and intersectional literature of multiple stigmatization. *Sex Roles*, 73(9), 408–413. <https://doi.org/10.1007/s11199-015-0543-4>
- Reno, R. R., Cialdini, R. B., & Kallgren, C. A. (1993). The transsituational influence of social norms. *Journal of Personality and Social Psychology*, 64(1), 104–112. <https://doi.org/10.1037/0022-3514.64.1.104>
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331–363. <https://doi.org/10.1037/1089-2680.7.4.331>
- Richman, L., & Leary, M. (2009). Reactions to discrimination, stigmatization, ostracism, and other forms of interpersonal rejection: A multimotive model. *Psychological Review*, 116(2), 365–383. <https://doi.org/10.1037/a0015250>
- Rowling, J. K. (2016). As if we needed more reasons to worship @andy_murray: He just reminded John Inverdale that women are people, too. Twitter. Retrieved from: https://twitter.com/jk_rowling/status/765129586489327618?ref_src=twsrc%5Etfw
- 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>
- Shelton, J. N., Richeson, J. A., Salvatore, J., & Hill, D. M. (2006). Silence is not golden: The intrapersonal consequences of not confronting prejudice. In S. Levin & C. V. Laar (Eds.), *Stigma and group inequality: Social psychological perspectives* (pp. 65–82). Mahwah, NJ: Lawrence Erlbaum Associates.
- Simon, B., & Klandermans, P. G. (2001). Toward a social psychological analysis of politicized collective identity: Conceptualization, antecedents, and consequences. *American Psychologist*, 56, 319–331. <https://doi.org/10.1037/0003-066x.56.4.319>
- Smith, J. R., & Louis, W. R. (2008). Do as we say and as we do: The interplay of descriptive and injunctive group norms in the attitude-behaviour relationship. *British Journal of Social Psychology*, 47(4), 647–666. <https://doi.org/10.1348/014466607X269748>
- Swim, J. K., & Hyers, L. L. (1999). Excuse me—what did you just say?! Women's public and private responses to sexist remarks. *Journal of Experimental Social Psychology*, 35(1), 68–88. <https://doi.org/10.1006/jesp.1998.1370>
- Swim, J. K., Hyers, L. L., Cohen, L. L., Fitzgerald, D. C., & Bylsma, W. H. (2003). African American college students' experiences with everyday racism: Characteristics and responses to these incidents. *Journal of Black Psychology*, 29(1), 38–67. <https://doi.org/10.1177/0095798402239228>
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Hall Publishers.
- TurkPrime (2018). *After the bot scare: Understanding what's been happening with data collection on MTurk and how to stop it*. Retrieved from <https://blog.turkprime.com/after-the-bot-scare-understanding-whats-been-happening-with-data-collection-on-mturk-and-how-to-stop-it>
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96. <https://doi.org/10.1037/0022-3514.92.1.82>
- Warner, R. H., & Branscombe, N. R. (2012). Observer perceptions of moral obligations in groups with a history of victimization. *Personality and Social Psychology Bulletin*, 38(7), 882–894. <https://doi.org/10.1177/0146167212439212>
- Warner, R. H., Wohl, M. J. A., & Branscombe, N. R. (2014). When do victim group members feel a moral obligation to help suffering others? *European Journal of Social Psychology*, 44(3), 231–241. <https://doi.org/10.1002/ejsp.2010>
- Wiley, S., & Dunne, C. (2019). Comrades in the struggle? Feminist women prefer male allies who offer autonomy- not dependency-oriented help. *Sex Roles*, 80(11), 656–666. <https://doi.org/10.1007/s11199-018-0970-0>
- Wilt, J., & Revelle, W. (2017). Extraversion. In T. A. Widiger (Ed.), *Oxford library of psychology. The Oxford handbook of the Five Factor Model* (pp. 57–81). Oxford, UK: Oxford University Press.
- Woodzicka, J. A., & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues*, 57(1), 15–30. <https://doi.org/10.1111/0022-4537.00199>
- Woodzicka, J. A., Mallett, R. K., Hendricks, S., & Pruitt, A. V. (2015). It's just a (sexist) joke: Comparing reactions to sexist versus racist communications. *HUMOR: International Journal of Humor Research*, 28(2), 25. <https://doi.org/10.1515/humor-2015-0025>
- Wout, D. A., Murphy, M. C., & Barnett, S. (2014). When having black friends isn't enough: How threat cues undermine safety cues in friendship formation. *Social Psychological and Personality Science*, 5(7), 844–851. <https://doi.org/10.1177/1948550614535820>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: HildebrandLK, Jusuf CC, Monteith MJ. Ally confrontations as identity-safety cues for marginalized individuals. *Eur J Soc Psychol*. 2020;00:1–16. <https://doi.org/10.1002/ejsp.2692>