

You(r behaviors) are racist: Responses to prejudice confrontations depend on confrontation focus

ACCEPTED VERSION

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's [AM terms of use](#), but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://doi.org/10.1007/s10869-022-09811-5>

Abstract

Directly confronting others' prejudice has been promoted as an important component in creating an inclusive workplace and motivating individuals to enact more inclusive behaviors. However, confrontations can also result in negative reactions from the confronted such as hostility or withdrawal. Across three studies (one experiment focused on anti-Black racism and two survey studies in which retrospective incidents of any type of racism were collected), we examine the extent to which the focus of a confrontation is associated with immediate constructive (i.e., prosocial) or destructive (i.e., antisocial, withdrawal/avoidant) responses. Specifically, we examine the differential effects of confronting racism in a way that focuses on the specific, temporary, and external behaviors enacted by the individual (behavior-focused) compared to focusing on the global, stable, and internal attributes of the person (person-focused). Across three studies we find converging evidence that behavior-focused confrontations are associated with more prosocial responses (e.g., self-improvement plans, appreciation) and that person-focused confrontations are associated with more antisocial and withdrawal/avoidant responses (e.g., hostility, avoidance). In Study 3, we find that interpersonal motivations (i.e., benevolence, revenge) and relationship repair efficacy help explain the effects of confrontation focus. We also find that the race of the confronter was not a significant moderator of findings, but we see some differences when comparing confronting anti-Black racism to confronting other types of racism (Studies 2 and 3). These results suggest that confrontation focus, and associated interpersonal motivations and beliefs, are important to understanding responses to prejudice confrontations. Implications are discussed.

**You(r behaviors) are racist: Responses to prejudice confrontations depend on
confrontation focus**

Recent popular books on anti-racism discuss the need to call out prejudice in others and avoid engaging in “White silence” (Kendi, 2019; Oluo, 2019; Saad, 2020). In organizations, confronting prejudice can be an important strategy for creating an inclusive climate (Ashburn-Nardo et al., 2008). However, some people react negatively to being confronted by others concerning their biases, becoming defensive or hostile (Czopp et al., 2006). These negative responses may undercut the effectiveness of prejudice confrontations as a tactic for promoting workplace harmony, and may even result in more strained relations in the workplace. Hence, we focus here on the qualities of confrontations concerning racism that determine both constructive and destructive responses.

Past findings suggest that more “extreme” or “hot” confrontations concerning racism can lead to a more negative response (Czopp et al., 2006) or negative evaluation of the confronter (Schultz & Maddox, 2013) than less extreme or “milder” confrontations, differentiating confrontations by the perceived emotionality of the language used or overall tone. We extend this prior work on the content of confrontations by examining who or what is characterized as problematic in the confrontation (i.e., the focus of the confrontation). Negative feedback can be communicated as more of a reflection of stable and global qualities of the person or more of a reflection of temporary and changeable behaviors (Abramson et al., 1978; Peterson et al., 1982). This distinction may be important to understanding outcomes of race-based confrontations, as feeling negatively about one’s self can lead to hopelessness and disengagement (e.g., Hui et al., 2012; Joiner, 2001), whereas feeling negatively about things one can change can lead to persistence in the face of adversity (e.g., Foll et al., 2006; Ilgen & Davis, 2000).

Some evidence suggests that negative feedback targeting personal and/or unchangeable qualities results in negative ratings of those giving the negative feedback (Liden et al., 1988; Waung & Jones, 2006). However, prior research has not examined whether the attributions communicated in this feedback impacts recipients' interpersonal behavior toward the individual providing the feedback. Furthermore, these dynamics have not, to our knowledge, been examined within the context of prejudice confrontations. We contribute to the literature by testing whether the attributions implied by prejudice confrontations determine whether recipients respond in a positive or negative manner. We examine this main relationship across three different studies: one experiment (Study 1) and two retrospective account studies (Studies 2 and 3), the latter of which focuses specifically on workplace confrontations. In Study 1 we focus on anti-Black racism confrontations. Studies 2 and 3 include racism confrontations concerning any racial group, allowing for comparisons that can illuminate important differences between different types of racial confrontations. Beyond focusing specifically on workplace confrontations, we further extend beyond the previous studies in Study 3 by examining the cognitions and motivations that may underlie the relationship between confrontation focus and subsequent responses.

Examining these dynamics in the context of race-based confrontations is an important contribution to the literature. Racism confrontations need to be studied explicitly, as they differ from the types of feedback traditionally studied (e.g., task, overall work) in important ways. Specifically, racism confrontations are more likely to be connected to one's morality or character (Sommers & Norton, 2006) rather than one's competence or skill. Morality is highly influential in global evaluations (Brambilla & Leach, 2014; Goodwin et al., 2014), and, consequently, people tend to respond to threats to their morality with defensiveness (Wenzel et al., 2020).

Thus, it is possible that even a behavior-focused confrontation might fail to elicit constructive responses. Further, the nature of this type of confrontation makes the race of the confronter particularly relevant in terms of responses to the confrontation (Czopp et al., 2006; Winslow, 2004), compared to other types of confrontations. In Studies 2 and 3 we assessed whether reactions are amplified when being confronted by someone from the group they have targeted (e.g., a Black person confronting someone about their anti-Black racism). By examining the type of feedback given in a race-based confrontation and the racial characteristics of the parties involved, our findings can better assess any barriers unique to this type of confrontation that prevent constructive responses to feedback around a topic on which many approach with deep apprehension.

We further contribute to understanding race-based confrontations by examining potential mediating mechanisms in Study 3. Based in research on interpersonal responses to rejection (Smart Richman & Leary, 2009) as well as literature focused on responses to relational transgressions and offenses (e.g., Bradfield & Aquino, 1999; McCullough et al., 1998), we test the extent to which confrontation focus leads to differing cognitions and motivations pertaining to the relationship between the confronter and the confronted, which may mediate the effects of confrontation focus on recipients' reactions to confrontations. Understanding the mechanisms behind the effects of confrontation focus is key to promoting productive responses to confrontations. Specifically, examining mediators moves us beyond focusing on how the message can be delivered differently and points to ways in which organizations and individuals can design useful interventions for reducing the likelihood of an individual reacting poorly when confronted.

Lastly, our work specifically contributes to this special issue on understanding anti-Black racism in organizations. Racist incidents at work ranging from subtle microaggressions to blatant acts of racism add significant stress to the lives of Black employees (Clark, Anderson et al., 1999; Pitcan et al., 2018; Sue et al., 2008). We view confrontation as one key way to reduce negative experiences for Black employees, through both making individuals aware of their hurtful behavior, as well as setting a clear boundary as to what is (un)acceptable behavior in the workplace (Ashburn-Nardo et al., 2008). As such, it is important to understand what types of confrontations may lead to constructive versus destructive responses. Further, Black employees in the United States (our research context) have a specific experience and history of discrimination (Reskin, 2012), and Black professionals are subjected to specific negative stereotypes (e.g., Fiske et al., 2002). As such, we aim in these studies to both specifically study confrontations of anti-Black racism (Study 1) as well as explore differences between anti-Black racism and other types of racism in the confrontation context (Studies 2 and 3).

Background

Interpersonal Reactions to Confrontations

Race-based confrontations may be effective because they facilitate learning about one's biases. Indeed, being made aware of one's prejudices has been connected to constructive, positive responses, such as engaging in efforts to reduce one's prejudiced attitudes (Amodio et al., 2007; Monteith et al., 2010). However, confrontations concerning prejudice can lead to negative behaviors that are interpersonally destructive, such as hostility and denial (Czopp et al., 2006). In research specifically examining race-based confrontations, aspects of the confrontation message appear influential in determining whether the confrontation elicits positive or negative responses. For example, Czopp and colleagues (2006) compared reactions to "hot" and "cold"

confrontations, differentiated in terms of the degree that the confrontation was personally threatening versus casual and non-threatening (respectively). The “hot” confrontation involved describing the person and their actions with the terms *racist* and *prejudiced*, whereas the “cold” confrontation suggested the individual should be *a little more fair*. They found that when confrontations were more threatening, participants were less likely to accept the confrontation and more likely to view the confronter negatively. Schultz and Maddox (2013) found when a Black college student made an “extreme” claim of racial bias, operationalized as using clear and forceful language that discussed biases against Black students and explicitly discussed how the campus was majority White, he was viewed more negatively than when he used “milder” language that suggested Black students have a more difficult time on campus without explicitly mentioning that White people are in the majority. These findings adeptly highlight that different types of confrontations will yield different types of reactions. In the current research, we extend this approach by examining the role of feedback focus in guiding interpersonal responses to race-based confrontations, building from Smart Richman and Leary’s (2009) multi-motive theory of interpersonal rejection.

When someone confronts another to tell them they are prejudiced or that they did/said something prejudiced, it likely signals to that person that who they are or what they did is not acceptable. Indeed, people associate prejudice, and specifically racism, with negative personal qualities such as hatefulness, lacking of empathy, and closed-mindedness (Sommers & Norton, 2006). Thus, this type of confrontation may threaten the confronted person’s relational value (i.e., the degree to which they are viewed as desirable for belonging in relationships and groups), akin to experiences of interpersonal rejection (Smart Richman & Leary, 2009). Smart Richman and Leary’s multi-motive model posits that people exhibit three primary motivated responses

when their relational value is threatened: prosocial responses, antisocial responses, and withdrawal/avoidance responses. Prosocial responses include those aimed toward gaining interpersonal acceptance and repairing the relationship, such as expressing affiliation, cooperation, engagement, and agreement. Antisocial responses involve defending oneself against the rejection or lashing out at the person who rejected them. Lastly, withdrawal/avoidant responses involve efforts to avoid further pain by removing oneself from the situation through physical avoidance or psychological disengagement.

According to this multi-motive model (Smart Richman & Leary, 2009), when there is little expectation that one can repair interpersonal acceptance following a rejection, the individual is more likely to engage in antisocial and withdrawal/avoidant responses and less likely to engage in prosocial responses. Race-based confrontations focusing more on the stable, global, and internal attributes of a *person's being* should lead to more destructive responses (i.e., antisocial, withdrawal/avoidant) and confrontations that focus more on the temporary, specific, and external attributes of a *person's behavior* should lead to more constructive or prosocial responses. For parsimony, we will label these two different types of confrontation feedback as *person-focused* (i.e., internal, stable, global) or *behavior-focused* (i.e., external, temporary, specific) from this point forth.

Race-based confrontations suggesting the person's character is problematic may lower expectations that the situation or relationship can be repaired, and thus should elicit unproductive responses, whereas confrontations emphasizing changeable and circumscribed behaviors should be more likely to elicit constructive responses because they foster optimism that the situation or relationship can be improved. Research on how individuals explain their own failures to themselves (i.e., their causal attribution styles, Abramson et al., 1978) gives further insight into

the importance of this distinction. Specifically, three types of causal attributions have been grouped together frequently as representing psychologically healthier versus unhealthier attributional styles (Abramson et al., 1978; 1989; Fresco et al., 2006; Peterson & Barrett, 1987; Joiner, 2001; Anderson et al., 1994; Hu et al., 2015; Houston, 2016; Sanjuán et al., 2008). Events can be attributed to qualities of the person (internal) or situation (external), to causes that are transient (unstable/temporary) or enduring (stable), and to causes that are present across many situations (global) or circumscribed to particular situations (specific; Abramson et al., 1978; Peterson et al., 1982).

Several studies suggest that people are more likely to persist on tasks or goals when they attribute setbacks to external, temporary, and specific causes (e.g., Andrews & Debus, 1978; Foll et al., 2006; Graham, 1991; Ilgen & Davis, 2000; Nauta et al., 1999). Furthermore, this attributional pattern appears to buffer the experience of negative affect following failure (Graham, 1991; McFarland & Ross, 1982; Metalsky et al., 1982). Particularly relevant to expectations of relationship repair, negative attributional patterns have also been linked to learned helplessness and distress (Hu et al., 2015; Joiner, 2000; Sanjuán et al., 2008). Thus, it is likely that being told one is racist will lead one to be less hopeful about the prospects of repairing the relationship than being told one did or said something racist.

It is important to note that we are focusing on a specific type of attributional style, but attributions can vary on other types of dimensions beyond the three we discussed, such as the degree of perceived controllability (Weiner, 1985) and intentionality (McGraw, 1987). Although we felt controllability (i.e., having control over one's actions) was not particularly relevant in the case of saying/doing something racist and intentionality was less directly linked to expectations one could repair the situation, events that are attributed to global, stable, and internal causes do

also tend to be seen as more controllable and more intentional (Anderson, 1983). Overall, we argue that confrontations suggesting that one's racial insensitivity stems from one's stable and global personality or character will lead to less constructive responses than getting feedback that one's racial insensitivity stems from one's temporary and specific behaviors.

Hypothesis 1: Behavior-focused confrontations will be associated with confrontation recipients' greater prosocial responding

Hypothesis 2: Person-focused responses will be associated with confrontation recipients' greater antisocial and withdrawal/avoidant responding

Although our main aim is to study the effects of confrontation focus on confrontation responses, it is also useful to test the overall effect of being confronted relative to not being confronted (i.e., a control condition). There is reason to believe that any type of confrontation will elicit negative responses relative to not being confronted, given that confrontations imply negative evaluations. Findings from two of Czopp and colleagues' (Studies 2 and 3, 2006) studies that included a control condition support this proposition. Specifically, they found that compared to a no-confrontation condition, individuals who were confronted relatively mildly (e.g., "some of your answers seemed a little offensive") still felt worse about themselves. Thus, we predict that person- and behavior-focused confrontations will be associated with greater destructive responses compared to not being confronted.

Hypothesis 3: Person- and behavior-focused confrontations will be associated with confrontation recipients' greater destructive responses, relative to a no-confrontation control condition

Shame and Guilt

In addition to interpersonal responses, we also aimed to examine differences in emotional reactions to being confronted. Specifically, Tangney and colleagues' (1992; 1996) differentiation of shame and guilt emotions are directly tied to person- versus behavior-focused distinctions. Tangney and colleagues (1992; 1996) argue that shame and guilt are both negative feelings, yet they differ as to their "object of concern" (1992, p. 169). With shame, one feels bad about the self, marked by perceptions of deficiency (e.g., *I am a bad person*). With guilt, the worry is not about one's self but rather one's actions. As such, one can feel guilty about one's behavior without feeling like a bad person (e.g., *I did something bad*). Given this theoretical distinction, a behavior-focused confrontation may lead to more guilt, but less shame, than a person-focused confrontation. As all types of confrontations are more likely to elicit negative emotions than not being confronted at all, we also posit that both types of confrontations will lead to more guilt and shame than not being confronted. Czopp and colleagues (2006) tested the effect of race-based confrontations on negative emotions toward oneself, but grouped together shame- and guilt-related feelings. They found no difference between low and high threat confrontation conditions in negative feelings toward oneself, but did find more negative feelings when being confronted versus a control condition. However, based on the theoretical distinctions between shame and guilt, we predict we will also find differences between person- and behavior-focused conditions on these emotional responses.

Hypothesis 4a: Person- and behavior-focused confrontations will be associated with confrontation recipients' greater shame and guilt relative to a no-confrontation control condition

Hypothesis 4b: Person-focused confrontations will be associated with confrontation recipients' greater shame relative to behavior-focused confrontations

Hypothesis 4c: Behavior-focused confrontations will be associated with confrontation recipients' greater guilt, relative to person-focused confrontations

Overview of Studies

Across three studies, we examine the role of confrontation focus -- whether confrontations focus on specific problematic behaviors or the people enacting those behaviors -- in eliciting different types of reactions to a race-based confrontation (H1 and H2). In the first study, we use an experimental paradigm (including a control condition) focused on confronting anti-Black racism and also examine differences in responses to confrontation versus no confrontation (H3) and differences in emotional responses (H4). In the second study, we gather responses to retrospective incidents of confronting racism toward any racial group, examining the role of confronter race (H5) and exploring the role of the group targeted by the racist action. In the third study, we also collect retrospective incidents, restricted to those that occurred while at work. We extend beyond Study 2 by examining interpersonal motivations and cognitions as potential mediators of the relationship between confrontation focus and response (H6 and H7).

Study 1

Participants

Data were collected from 90 undergraduate students (69 female, mean age = 19.2; SD = 2.01) at a large Mid-Atlantic university. Participants received either course credit (1 credit/hour) or monetary compensation for their participation (\$10). Individuals who participated for course credit were students enrolled in undergraduate psychology courses. Participants were 52.2% White, 31.1% Asian, 6.7% Latino and 10.0% mixed-race. Black participants were excluded from this experimental paradigm because the interest was in understanding reactions to confrontations of racism toward Black individuals by non-Black individuals. As we did not want

to clue participants in to the true nature of the study, we did not advertise the fact that Black participants would be excluded. Thus, 15 Black participants (not included in the original count of 90 participants) took part in the control condition only and were excluded from data analysis. Seven additional participants were excluded from all analyses for failing to pass instructed response attention check items (e.g., “please choose Strongly Agree for this item”) embedded in the survey (2 from the control condition, 4 from the behavior-focused condition, 1 from the person-focused condition), and 10 additional participants were excluded for expressing suspicions as to the purpose of the study (4 participants in the behavior-focused condition, 6 participants in the person-focused condition) leaving a final sample size of 73 (81.1 % of original sample). Participants were conservatively probed for suspicion. During debriefing, participants were told that the true nature of the study was not exactly what it had been described as and were asked if there was anything else they could think of that the study was investigating. If a participant expressed anything related to race or racial confrontations, they were excluded for suspicion. Exclusion of participants from data analysis did not change the significance or direction of results.

Procedure

One participant per session reported to the laboratory. Participants were placed into one of three conditions upon beginning the study via random assignment: person-focused confrontation, behavior-focused confrontation, or no confrontation (control). All participants were told that the study would involve interacting with another participant and that the researchers sought to better understand the process of online communication in the workplace.

The study began with the participant and their study partner (a White confederate who was naive to condition) completing a task in which they looked at photographs of people with a

short phrase underneath (e.g., works with children) and then were asked to each make a different inference about that individual (e.g., she's a pediatrician; she's a kindergarten teacher). The participant and confederate alternated as to who went first in describing the person in the photo and who had to follow with an inference different from the first answer given. The task primed participants to make stereotypical statements about Black individuals, such that some of the photos were of Black individuals and the attached phrase (e.g., is found behind bars) prompted most participants to make a stereotypical response (e.g., he's a criminal), followed by the confederate who always made a non-stereotypical response (e.g., he's a bartender). The task was arranged so that the participant would always go first for the stereotypical photo/phrase combinations (see Czopp et al., 2006 for a complete overview of the task). After completing this task, the participant and confederate were separated. Participants assigned to the confrontation conditions were instructed to send their study partner an electronic message providing them with feedback on how they thought their partner did during the picture task. After the participant sent this message, they received a reply message from their partner that contained a confrontation. For participants in the control condition, no emails were exchanged. After reading this email the participant completed measures of their current emotional state, as well as evaluations of their study partner. In order to maintain the cover story, participants also completed a questionnaire assessing their attitudes toward online communication (Ledbetter, 2009). After completing all measures participants were then instructed to send another message to their partner, which was coded using the procedures described below. Participants were then thoroughly debriefed and probed for any suspicions as to the nature of the study.

Materials

Confrontation Manipulation

Participants were presented with one of two possible confrontation manipulations or were not presented with any confrontation at all. Participants randomly assigned to the *behavior-focused confrontation* received the following text, “Your answers were really negative toward Black people during that exercise. I’m sure you don’t normally make insensitive comments and you seem like someone who respects others in general, but those answers could really harm others.” Participants randomly assigned to the *person-focused confrontation* received the following text, “You were really negative toward Black people during that exercise. You seem like someone who is racist toward Black people in general and probably has been for a long time. You could really harm others with your attitudes.”

We expect that the behavior-focused confrontation conveys that racial prejudice is specific to a behavioral act that is unstable over time and external to the person, whereas the person-focused confrontation conveys that racial prejudice is a stable feature of the person across time and situations. To ensure that our conditions vary on these dimensions and are differentiated from previous distinctions of high versus low threat confrontations (Czopp et al., 2006)¹, we conducted a pilot study. As the scenario was concerning reactions to confrontations of racism toward Black individuals, only non-Black participants completed the pilot study. Seventy-six college student volunteers (78.9 % women; 17.1 % Asian, 71.1 % Caucasian/White, 7.9 % Hispanic/Latinx; 2.6 % Middle Eastern, 1.3 % Multi-racial) imagined that they had just discussed Black-White race relations with another individual and that they had said something insensitive during that interaction. Participants were then asked to imagine they were confronted by this individual and asked to rate the attributes of the confrontation manipulation. Specifically, participants indicated the extent to which they perceived the confronter’s comments reflected a belief that the participant’s actions were due to circumstances (external attribution) or due to

themselves (internal attribution) on a scale of 1 (*totally due to circumstances*) to 7 (*totally due to me*); that the participant would not make the same types of comments again (unstable attribution) or would (stable attribution) on a scale of 1 (*never again*) to 7 (*always*); and that the insensitive comment they made reflected something about them that was specific to the situation (specific attribution) or is present in many situations in their life (global attribution) on a scale of 1 (*is particular/specific to the situation*) to 7 (*comes up in all situations in my life*). Pilot data means are presented in Table 1. Results suggest that the person-focused confrontation in this study conveyed significantly more stable, internal, and global attributions relative to the behavior-focused confrontation [$t(75) = 19.17, p < .001$; $t(75) = 18.00, p < .001$; $t(74) = 16.65, p < .001$; respectively] and also relative to the high threat condition used by Czopp and colleagues [$t(74) = 10.53, p < .001$; $t(74) = 8.39, p < .001$; $t(74) = 7.97, p < .001$; respectively]. In addition, the behavior-focused condition elicited significantly more temporary, external, and specific attributions relative to the low threat condition used by Czopp and colleagues [$t(73) = -4.42, p < .001$; $t(73) = -3.00, p = .004$; $t(73) = -3.40, p = .001$; respectively]. Hence, these results suggest the validity of the behavior-focused and person-focused manipulations used in the current research, and suggest that these manipulations vary in behavior- and person-focus more than high threat and low threat manipulations used in prior research, which were not intended to vary along these dimensions.

Measures

Qualitative Coding

Participants' written responses to the confederate following the confrontation were coded by a team of objective observers. Behavioral categories include those used in prior research on prejudice confrontations that fit the definitions of prosocial (i.e., acknowledgement, bias

recognized, apology) or antisocial (i.e., denial, hostility) responses (Czopp et al., 2006) as well as two additional behaviors identified by the present research team that fit under the prosocial category (i.e., desire to self-improve, appreciation/thankfulness), resulting in a total of seven behaviors. No withdrawal/avoidant behaviors emerged from the data, likely due to the difficulty of expressing withdrawal or avoidance in a written reaction to a confrontation. We relied on a quantitative measure of the individual's willingness to work with the partner again as a measure of withdrawal/avoidant response (discussed below). Participant responses to the confrontation sent to them were coded for each of the seven behaviors.

Acknowledgement was defined as any form of verbal concession or understanding of the confronter's point of view, such as repeating what was said to them (e.g., "I know what you mean," "I understand"). *Bias recognized* was defined as expressing awareness of or admitting the possibility that their response could have been influenced by race and therefore biased (e.g., "Now I realize that I too make some judgments too quickly," "It's sad but they were the first thoughts that came to my head"). *Apology* was defined as expressing any sort of regret (e.g., "I apologize," "I'm sorry"). *Self-improvement* was defined as a willingness to accept new ideas or acquire new skills as a result of confrontations, with emphasis on changing oneself/attitudes (e.g., "I'm going to stop making these kinds of comments," "I want to change my attitudes"). *Appreciation/thankfulness* was defined as expressing gratitude for bringing prejudiced remarks to their attention (e.g., "Thank you for the feedback," "I appreciate your honesty").

Denial was defined as denying or rejecting the suggestion that the participant was a prejudiced person or exhibited prejudiced behavior. Denial was initially broken down into two subcategories, one focused on the self (e.g. "I'm not a racist") and one focused on behavior (e.g., "I did not display any type of racism"), however, these ratings were highly correlated and were

later combined for analyses. Examining them as separate subcategories does not change interpretations of results. *Hostility* was defined as conveying anger, resentment, or condescension toward the individual (e.g., “Don’t be so sensitive”).

Eight individuals, who were naive to confrontation focus condition, coded all qualitative responses. Participant responses were coded for the extent to which they displayed prosocial (acknowledgment, bias recognized, apology, self-improvement, appreciation/thankfulness) and antisocial (denial, hostility) behaviors. Each response was coded on a 1-7 Likert scale for the extent to which the response displayed prosocial or antisocial behaviors (1 = not at all, 7 = very much). We had coders (undergraduate research assistants) rate these behaviors as to the extent to which they were displayed rather than being present or not, warranting measurement of interrater consistency over interrater absolute agreement (McGraw & Wong, 1996). Interrater consistency was sufficient for each behavior ($ICC > .87$ for all behaviors) and thus codes were collapsed across the eight coders so that each response had a specific score indicating mean rating across all raters.

Additional Measures

We included one item asking participants if they would want to work with their partner again on a 5-point strongly disagree to strongly agree scale, as a measure of withdrawal/avoidant responding. Participants’ felt shame ($\alpha = .87$) and guilt ($\alpha = .89$) were assessed with four- and five-item measures (respectively) by Tangney and colleagues (1996), using a 5-point response scale (strongly disagree to strongly agree). Participants provided their gender, age, and race/ethnicity. We also included other variables that are not germane to our hypotheses. More information about these variables and any differences by condition are available in the Supplementary Materials to this manuscript.

Study 1 Results

See Table 2 for correlations between main study variables. As seen in Table 3, for the prosocial responses, appreciation/thankfulness was displayed to the greatest extent out of all prosocial behaviors, followed by acknowledgement. For antisocial behaviors, denial was displayed to a greater extent than hostility.

To test Hypotheses 1 and 2, the control condition was excluded from these specific analyses, as the comparison of interest is between person- and behavior-focused confrontations. Control variables included the number of negative statements made by the individual during the picture task (1, 2, or 3) and participant race (0 = White, 1 = non-Black POC). To test H1 and H2, which posited that person-focused confrontations will be met with less prosocial responses and more antisocial and withdrawal/avoidant responses when compared to behavior-focused confrontation, separate hierarchical linear regression analyses were conducted in which prosocial responses (i.e., the mean of acknowledgement, bias recognition, self-improvement, apology, and appreciation/thankfulness ratings), antisocial responses (i.e., the mean of denial and hostility responses), and the withdrawal measure (willingness to work with the individual again) were regressed on condition (behavior-focused = 0, person-focused = 1). We also conducted separate regressions for each specific prosocial and antisocial behavior (see Table 3). Power analyses for regression indicated we need 55 participants to detect a medium effect size and 25 participants to detect a large effect size with .80 power for a two-tailed test with one predictor variable and two control variables. With our exclusions, we had 46 participants (22 in the behavior-focused condition and 24 in the person-focused condition), meaning we are slightly underpowered to detect a medium effect and adequately powered for a large effect.

As seen in Table 3, participants in the behavior-focused confrontation enacted more prosocial responses overall ($b = -1.10, p = .003, 95\% \text{ CI } [-1.81, -.40], \Delta R^2 = .19$) compared to the person-focused condition and more of the following types of prosocial responses: acknowledgement of the confrontation ($b = -1.15, p = .008, 95\% \text{ CI } [-1.97, -.32], \Delta R^2 = .15$), recognition of one's own bias ($b = -1.40, p = .002, 95\% \text{ CI } [-2.27, -.53], \Delta R^2 = .20$), desire/plans to self-improve in regard to one's own prejudice ($b = -1.32, p = .010, 95\% \text{ CI } [-2.31, -.33], \Delta R^2 = .15$), and appreciation/thankfulness for the confrontation ($b = -1.66, p = .003, 95\% \text{ CI } [-2.73, -.58], \Delta R^2 = .18$). For example, someone in the behavior-focused condition expressed their appreciation/ thankfulness and self-improvement desires:

“I'm sorry if I at all offended you with my answers. I truly didn't mean to be insensitive in any way. I thought I was conscious of the way my answers portrayed African Americans, but in the future I will make an active effort to be more aware of the implications of my words. Again, I apologize for inflicting any harm. Thank you for letting me know about this.”

Also seen in Table 3, the person-focused confrontation resulted in more antisocial responses overall ($b = 1.35, p = .001, 95\% \text{ CI } [.56, 2.14], \Delta R^2 = .22$) and of the following specific antisocial responses: denial ($b = 1.71, p = .001, 95\% \text{ CI } [.72, 2.70], \Delta R^2 = .22$) and expressed hostility ($b = 1.12, p = .010, 95\% \text{ CI } [.28, 1.95], \Delta R^2 = .15$). As one example in the person-focused condition, this participant displayed both hostility toward the confronter and denial:

“Wow. That is kind of mean to say that. You do not know me, and you should not judge me based off a silly little exercise like that. I went the easy route and said the first things that came to mind, and they were probably the first things that came to your mind too. I

am not racist, and basing your judgement on that exercise is ridiculous. You seemed really nice in person, smiling and all, but the more I think about it, you are not.”

Overall, the results for the mean of prosocial and antisocial responses, as well as the individual ratings of acknowledgement, bias recognized, self-improvement, appreciation/thankfulness, denial, and hostility align with hypotheses. However, there were no significant differences between person- and behavior-focused confrontations in the prosocial response of apology nor the withdrawal response of willingness to work with the person again. In sum, both H1 and H2 were partially supported.

To test H3, which predicted that both confrontation conditions (person- and behavior-focused) would be associated with more withdrawal responses, willingness to work with confronter again was regressed on condition (person-focused versus behavior-focused versus control). Condition was dummy-coded with control being the referent variable. Power analyses for regression indicated we need 68 participants to detect a medium effect size with .80 power with two predictor variables and three control variables, which we exceeded ($N = 73$).

Participants were less likely to want to work with the confederate again if they had been in either confrontation condition, compared to control (behavior-focused: $b = -.57, p = .027, 95\% \text{ CI} [-1.07, -.07]$; person-focused: $b = -.77, p = .003, 95\% \text{ CI} [-1.27, -.26]; R^2 = .12$). Overall, H3 was supported.

To test H4a-c, which predicted that confrontation conditions would affect feelings of shame and guilt, we conducted separate regressions in which shame or guilt were regressed on confrontation condition. To test the differences in experimental versus control conditions (H4a), we used control condition as the referent. Participants were more likely to report feeling shame (behavior-focused: $b = 1.03, p < .001, 95\% \text{ CI} [.49, 1.56]$; person-focused: $b = .66, p = .016, 95$

% CI [.13, 1.19]; $R^2 = .23$) and feeling guilt (behavior-focused: $b = .93$, $p = .002$, 95 % CI [.36, 1.51]; person-focused: $b = .58$, $p = .049$, 95 % CI [.003, 1.15]; $R^2 = .13$) in the confrontation conditions as compared to control, supporting H4a. To test the differences between confrontation conditions, we reran analyses using the behavior-focused condition as the referent. Participants did not differ between person- and behavior-focused conditions on either shame ($b = -.37$, $p = .190$, *ns*) or guilt ($b = -.36$, $p = .239$, *ns*). H4b and H4c were not supported.

Study 1 Discussion

Results from Study 1 generally support predictions that confrontations evoking temporary, external, and specific attributions (i.e., behavior-focused confrontations) will lead the confronted party to engage in more prosocial responses, whereas confrontations evoking stable, internal, and global attributions (i.e., person-focused confrontations) are more likely to result in denial and hostility. These findings support past research that found that the type of confrontation matters (Czopp et al., 2006), but extend beyond this work by highlighting the role of confrontation focus. Specifically, we find support for the positive effects of feedback that focuses on temporary, external, and specific causal attributions following a personal failure (Graham, 1991; McFarland & Ross, 1982; Metalsky et al., 1982). When made aware of their transgression while explicitly separating it from them as a person, participants were more likely to recognize their own biases, express intentions to change (i.e., self-improve), express appreciation for the feedback, and acknowledge the feedback. Overall, these findings are in line with aspects of the multi-motive model of interpersonal rejection, which posits that perceived reparability of the transgression is associated with the extent to which one engages in different types of responses (Smart Richman & Leary, 2009). Specifically, we theorized that person-focused confrontations are more likely to result in one behaving in ways that reflect a perceived

futility in trying to repair the situation and found that indeed, participants in this condition were more likely to deny and lash out at the confronter. In contrast, participants experiencing behavior-focused confrontations behaved more prosocially, suggesting they may have felt there was a path to redemption following a rejection incident.

Interestingly, we did not find a relationship between confrontation focus and our withdrawal response measure. Participants were less likely to want to work again with someone who confronted them (versus the control condition), but the type of confrontation did not affect this response. In Study 2, we ask if individuals actually engaged in withdrawal/avoidant responding following a confrontation, examining the extent to which these null findings might be due to the measure focusing on a desire to withdraw/avoid rather than actual withdrawal/avoidant behaviors.

Participants reported feeling more guilt and shame in both confrontation conditions compared to no confrontation, but findings did not support a distinction between shame and guilt feelings based on the focus of the confrontation. This is surprising as previous literature clearly connects internal-focused feedback to specific negative emotions (e.g., Hareli & Hess, 2008; Tangey et al., 1992; 1996; Weiner, 1986). Shame and guilt were highly correlated with one another ($r = .84$), meaning our participants may not have been making conceptual distinctions between these measures. It is also possible that confrontation focus directly changes beliefs surrounding the future of the relationship (Smart Richman & Leary, 2009) without differentiating emotional reactions. Future research directly examining relationship-related beliefs will elucidate this point.

There are limitations to this study that we attempt to address in Study 2. In Study 1, we used a laboratory paradigm, which improves our ability to establish causal effects of

confrontation focus, but is artificial in nature. Specific aspects of our manipulations (such as use of the term “racist”) may have been unrealistically direct, contributing to the effects. Further, our attempt to adhere to conceptual definitions of the different dimensions of person-focused (i.e., internal, stable, global) and behavior-focused (i.e., external, temporary, specific) resulted in two experimental conditions that were not completely parallel in language, bringing in other possible explanations for effects. Although our groupings of causal attributions are theorized to co-occur as different styles of explaining events (Abramson et al., 1978), we cannot say if one of the dimensions in particular was driving the effects. To address this limitation and increase external validity, in Study 2 we collect retrospective incidents of times people were confronted for saying/doing something racially insensitive, focusing specifically on if they perceived the confrontation as focused on their character/personality or their behaviors. This methodology allows us to analyze actual events and control for potential specific drivers of the effects (specifically, use of the term “racist”). Additionally, the confronter in Study 1 was always White and the target of the racism was always Black individuals. Past research suggests that the race of the confronter or target may affect perceptions of or reactions to the incident (e.g. Czopp et al., 2006; Inman & Baron, 1996). The retrospective approach of Study 2 allows us to examine and explore the race of the confronter and the target of the racism.

Study 2

In Study 2, we analyze retrospective accounts of individuals who have been subject to a race-based confrontation, examining the extent to which confrontation focus relates to different reactions to being confronted (H1 and H2). Retrospective accounts, while subject to memory distortions and other biases, have the advantage of allowing us to capture actual incidents of race-based confrontations, increasing our external validity. Further, as only one of the

confrontations in Study 1 used the term “racist”, we measured and controlled for use of the term in Study 2, allowing us to separate the effect of confrontation focus from the effects of using that specific and evocative term.

We also examined associations between the race of the confronter and responses to confrontations. Research on feedback and persuasion have noted that even with a strong argument, characteristics of the communicator are important predictors of how a person will react to and view a given message. For example, messages are best received when the communicator is viewed as credible, unbiased, and having constructive motives (Eagly et al., 1978; Esposito et al., 2013; Hovland et al., 1953; Steelman & Rutkowski, 2004; Wallace et al., 2020). In the context of race-based confrontations, race itself becomes a more salient communicator characteristic than in a typical negative feedback scenario. That is, being confronted by someone who is presumably personally offended and affected by one’s actions may elicit stronger interpersonal responses than being confronted by a non-targeted individual, because the former type of confrontation may more credibly communicate that one has caused harm to others. Czopp and colleagues (2003; 2006) have found mixed support for the effects of confronter race on reactions to the confronter, finding in one study that individuals felt worse about themselves when they were confronted by someone from the target group (i.e. a Black individual confronting about anti-Black racism; 2006), but found the opposite finding in a previous study (2003). However, it is important to note that only the former finding involved an actual (rather than hypothetical) interaction. As we are interested in understanding when confrontations can lead to constructive responses, it is important to clarify how confronter race may change the dynamics of the interaction and ultimately affect interpersonal responses.

There are several reasons why individuals might have stronger reactions to confrontations from target versus non-target individuals. For one, past research suggests that majority-group individuals are particularly concerned about appearing racist to target group members, compared to their own ingroup members (Winslow, 2004). Further, some have theorized and found that individuals pay more attention to the arguments of stigmatized individuals when they are motivated to appear nonprejudiced (Petty et al., 1999), suggesting that in a confrontation context, individuals may be more attuned to the focus of the confrontation when it is communicated by a targeted group member. Lastly, individuals tend to pay more attention to perceived experts (Heesacker et al., 1983; Tobin & Raymundo, 2009), and individuals may be perceived as more of an expert about racism toward their own group than toward another group. Thus, we predict that:

Hypothesis 5: Being confronted by someone belonging to the targeted racial group will lead to stronger reactions overall than being confronted by someone belonging to a different racial group, strengthening the relationship between confrontation focus and prosocial, antisocial, and withdrawal/avoidant responses.

We also explore any differences based on the race of the targeted group, so as to compare incidents of anti-Black racism (the focus of Study 1) with other types of racism. In general, organizational research examining race has tended to focus on one racial minority group or group together all BIPOC individuals in one group, comparing the experiences or perceptions of that group to those of White people (Ruggs et al., 2013). As such, previous work does not provide a solid basis for understanding the nuanced experiences of racism for different groups. For example, Black and Latinx groups are more likely to experience stereotyping related to inferiority compared to Asian groups, whereas Asian and Latinx groups are more likely to

experience stereotyping related to being foreign than Black groups (Zou & Cheryan, 2017). These different stereotypes may affect the perceived seriousness of the confrontation. Further, there is some evidence suggesting that the prototypical target of racism in the United States is a Black individual (Corning & Bucchianeri, 2010; Rodin et al., 1990), meaning that someone may feel less prepared to respond to race-based confrontations concerning other groups. The most recent available data from the FBI indicates that anti-Black hate crimes substantially outnumber reported hate crimes toward any other racial group in the United States (FBI, 2020). The unique discrimination and oppression experienced by Black Americans in the United States (Reskin, 2012) may make reactions to being confronted for anti-Black racism particularly strong, compared to other confrontations of racism. Overall, confronting racism could mean something different to the confronted individual, depending on the target of the racism. In this next study, we will explore any unique properties of confronting anti-Black racism.

Research Question: How do responses to confronting anti-Black racism differ from responses to confronting other types of racism?

Participants and Procedure

As we found no previous studies that collected descriptions of these types of incidents, we had no prior information as to the frequency of race-based confrontations. Thus, we asked potential participants if they were personally confronted about racial prejudice, but also cast a wider net and asked them to describe incidents in which they had been the *confronter* in a race-based confrontation or they had *witnessed* a race-based confrontation. Our main focus is on the perspective of someone who is confronted (to match the perspective of Study 1), but we also tested our predictions using retrospective accounts from the perspectives of confronters and witnesses.

We recruited participants via the online survey platform CINT. Thirteen hundred and thirty-one participants entered the survey, of which 768 were immediately prevented from entering the study due to not having experienced any of the three potential types of incidents we outlined. Of the 563 participants who responded affirmatively to having experienced at least one of the types of confrontation experiences, 235 reported having been confronted, provided an actual written response (filtering out those who said “none”, “N/A”, etc), and passed a thorough check to weed out automated “bot” responses (e.g., random series of letters, words, or numbers). These participants were the focus of our analyses. Three undergraduate research assistants read through every submitted account and flagged any incident where they suspected a “bot” response. One of the principal researchers then also read every response and made the final call on excluding a given case.

Data from participants who failed an instructed response attention check item were excluded from analyses ($N = 57$), as were data from participants who described incidents that were off topic (e.g., incidents about someone’s general feelings about racism and not about a specific incident; $N = 9$). We also excluded data from two participants who did not complete all of the measures included in our analyses. After these exclusions, we analyzed data from 169 participants (95 female, mean age = 43.8; $SD = 17.1$) who had experiences of being confronted (71.9 % of the original sample of those who had an incident of being confronted). We also analyzed data (using the same exclusionary criteria) from 192 participants who had experiences of being a confronter (72.2% of original sample) and 292 participants who had witnessed a confrontation (73.6 % of original sample), which is discussed in more detail in the Additional Analyses section below. Although these exclusions comprise a significant portion of our data, we followed best practices of incorporating attention check items (specifically *instructed response*

items), which “have the advantage of providing an obvious metric for scoring as correct or incorrect” (Meade & Craig, 2012, p. 452). Adopting these exclusion criteria increase the integrity of our data and the inferences made from our analyses. Table 4 provides descriptive statistics concerning the confronter and aspects of the confrontation.

Measures

Confrontation Focus

To measure confrontation focus, participants responded to two single-item measures on the extent to which the confrontation was person-focused (*The person confronting me focused on issues they had with my character or personality*) and on the extent to which the confrontation was behavior-focused (*The person confronting me focused on issues they had with my specific behavior and/or words*). Both items used a 1 (*Strongly Disagree*) to 6 (*Strongly Agree*) response scale.

Confrontation Responses

To measure confrontation responses, we generated items that corresponded to prosocial, antisocial, and withdrawal/avoidant responses, informed by theory (Smart Richman and Leary, 2009) and matched to our previous measures in Study 1. Specifically, five items measured prosocial responses (*Following the confrontation, I: apologized to the person; told the person I understood their point of view; told the person that I recognized what I did/said may have been wrong; told the person that I appreciated their feedback; told the person I would try not to do/say something like that again*), three items measured antisocial responses (*Following the confrontation, I: became angry with the person; denied the person’s accusations; yelled at or became aggressive with the person*) and two items measured withdrawal/avoidant responses (*Following the confrontation, I: tried to distance myself from the person; tried to end the*

conversation). All items used a 1 (*Strongly Disagree*) to 6 (*Strongly Agree*) response scale. An exploratory principal components analysis using varimax rotation revealed two dominant factors (explaining 67.9 % of the variance), with antisocial and withdrawal/avoidant loading on the same factor. However, confirmatory factor analyses of these scales support a three-factor structure over a two-factor structure [$\Delta\chi^2(2) = 12.34, p = .002$; CFI = .94, RMSEA = .10, SRMR = .08], with antisocial and withdrawal/avoidance remaining separate. Due to the superior model fit and the conceptual differences between antisocial and withdrawal/avoidant responding, we chose to create three total scales: a prosocial response scale ($\alpha = .93$), an antisocial scale ($\alpha = .69$), and a withdrawal/avoidance scale ($\alpha = .75$).

Additional Measures

Participants were asked several other questions about the confrontation, including: how well they knew the confronter at the time of the confrontation (*not well at all, somewhat well, moderately well, or very well*), the confronter's race/ethnicity, the race/ethnicity of the group which the confronter thought the participant was being racially insensitive, and whether or not the confronter used the term "racist" during the confrontation (0 = no, 1 = yes). Participants were also asked to provide their own race/ethnicity, their gender, their age, their typical working hours, and the industry category of their job.

Study 2 Results

Main Analyses

Table 5 provides the correlations of all relevant variables. To test H1 and H2, separate hierarchical linear regression analyses were conducted in which either the prosocial or antisocial/avoidance response scales were regressed on the two confrontation focus items (*person-focused* and *behavior-focused*). In both equations, whether or not the term "racist" was used during the

confrontation (0 = *no*, 1 = *yes*) was entered as a control variable. Power analyses for regression indicated we exceeded the 64 participants we need to detect a medium effect size with .80 power with two predictor variables and one control variable. As seen in Table 6, participants reported engaging in prosocial responses to a greater extent when they reported that the confrontation was more focused on their behaviors ($b = .19, p = .031$). The extent to which participants reported that the confrontation focus was person-focused was not related to extent of prosocial response ($b = .03, p = .731, ns$). For antisocial responses, the opposite pattern emerged. The extent to which participants reported that the confrontation was person-focused was related to engaging in antisocial responding ($b = .20, p < .001$), whereas there was no such relationship for behavior-focused confrontations ($b = .01, p = .928, ns$). We found a similar pattern for withdrawal/avoidant responses, with person-focused confrontation relating to greater withdrawal/avoidant responding ($b = .30, p < .001$) and no relationship between behavior focus and withdrawal/avoidant responding ($b = -.01, p = .903, ns$). Of note, these effects were found while controlling for the confronter's use of the term "racist" during the interaction, which showed a significant main effect on both antisocial ($b = .59, p = .007$) and withdrawal/avoidant ($b = .67, p = .011$) responses. Further, the addition of confrontation focus items increased the amount of variance explained in these responses significantly beyond the use of the term "racist" ($\Delta R^2 = .07, p = .002$; $\Delta R^2 = .10, p < .001$; respectively). Overall, these results provide support for both H1 and H2.

H4 predicted that confrontations in which the confronter was of the same race as the target of prejudice (e.g., a Black individual confronting anti-Black racism) would result in stronger relationships between focus and response than confrontations in which the confronter was of a different race than the target of prejudice. To test this, we conducted three separate

hierarchical linear regressions, in which prosocial, antisocial, and withdrawal/avoidant responses were regressed on the use of the term “racist” (control variable), followed by the two confrontation focus items, followed by the similarity of the confronter’s race to the target’s race (0 = confronter’s race different from that of targeted race, 1 = confronter’s race is the same as that of the targeted race), followed by the product terms to test the interaction between racial similarity and confrontation focus. Power analyses for regression indicated we exceeded the 68 participants we need to detect a medium effect size with .80 power with two predictor variables and six total predictor and control variables. We found no significant interactions between confronter-target race similarity and confrontation focus, failing to support H4.

Exploratory Analyses

Study 2 differed from Study 1 in that it included other types of racism besides anti-Black racism. We wanted to explore if type of racism had any effects on our variables of interest. As seen in Table 4, in the majority of our collected incidents (56.3 %), Black individuals were the target of the racism that was being confronted. We compared anti-Black racism to all other types of racism on both outcome variables and both confrontation focus variables. We found that compared to all other types of racist incidents ($M = 4.04$, $SD = 1.74$), anti-Black racist incidents were viewed as less person-focused ($M = 3.26$, $SD = 1.86$). There were no other differences in main study variables. We also explored if target of prejudice (anti-Black racism versus other types of racism) interacted with confrontation focus variables to predict confrontation responses. Specifically, we regressed the outcome variable (prosocial, antisocial, or withdrawal/avoidant) on the control variable (used the term “racist”), followed by both confrontation focus variables and the targeted race variable (anti-Black versus other racism), followed by the interaction between each confrontation focus variable and the targeted race variable. The targeted race of the

incident did not interact with either confrontation focus variable to predict prosocial or antisocial responses. However, the targeted race variable had a significant interaction with behavioral focus to predict withdrawal /avoidant responses (see Figure 1). Specifically, simple slopes tests revealed that the extent to which the confrontation was viewed as behavior-focused was associated with reduced withdrawal/avoidant responding for other types of racism ($b = -.42, p = .005$), but not for anti-Black racism ($b = .18, p = .089, ns$). That is, for all types of racism, person-focused confrontations are associated with more antisocial and withdrawal/avoidant responding and behavior-focused confrontations are associated with more prosocial responding. However, behavior-focused confrontations are only associated with less withdrawal/avoidant responding when the target of the racism is not Black.

Additional Analyses – Other Perspectives

As previously stated, it was unknown (based on a lack of research) if we would be able to collect enough incidents of being confronted about racism from a general population of survey responders. As such, we also collected additional data on times that the participant was a confronter and times that the participant witnessed a confrontation. Specifically, we analyzed data from 192 participants (130 female, mean age = 41.8; SD =16.3) who had experiences of being a confronter (72.2% of original sample) and 292 participants (186 female, mean age = 41.5; SD =16.2) who had witnessed a confrontation (73.6 % of original sample). All items were similar to those used in the main analyses, with rewordings to convey the specific perspective of being a confronter or being a witness. The response items were similarly combined for prosocial responses (confronter perspective $\alpha = .95$; witness perspective $\alpha = .92$), antisocial responses (confronter perspective $\alpha = .75$; witness perspective $\alpha = .61$), and withdrawal/avoidant responses (confronter perspective $\alpha = .67$; witness perspective $\alpha = .62$). Note that the antisocial and

withdrawal/avoidant scales for the witness perspective have fairly low reliabilities and results should be interpreted cautiously.

For both those who were confronters and those who witnessed a confrontation (controlling for use of the term “racist”), more person-focused confrontations were associated with reports of more antisocial responses ($b = .18, p = .004$; $b = .11, p = .030$; respectively). Person-focused responses were not associated with withdrawal/avoidant responding for the confronter perspective incidents ($b = .10, p = .137, ns$), but there was a significant association for the witness perspective ($b = .15, p = .008$). There was no relationship for either perspective (confronter or witness) between reports of more behavior-focused confrontations and antisocial responses ($b = .05, p = .638, ns$; $b = .05, p = .393, ns$; respectively) or withdrawal/avoidant responses ($b = .00, p = .984, ns$; $b = .01, p = .859, ns$; respectively). Antisocial responding results mirror the findings of the main analyses but withdrawal/avoidant responding results only partially replicate main analysis findings for the witness perspective. Results differed, however, when analyzing prosocial responses. For those who recalled an incident when they were the confronter, neither person- ($b = -.03, p = .674, ns$) nor behavior-focused ($b = -.11, p = .274, ns$) confrontations related to prosocial responses. For those who witnessed a confrontation, only person-focused confrontations related to prosocial responses ($b = .15, p = .014$), and in an unexpected direction. Behavior-focused confrontations were not significantly related to prosocial responses, though in a direction consistent with hypotheses ($b = .11, p = .114, ns$).

Study 2 Discussion

Individuals in a situation where they were confronted with doing/ saying something racially prejudiced reported being more prosocial when the focus was on the behavior and more antisocial and withdrawal/avoidant when the focus was on them as a person. This aligns with

experimental findings from Study 1 on reactions to being confronted. These relationships held outside of an artificial and controlled laboratory study and held when we controlled for using the term “racist”, suggesting that our Study 1 findings were not merely driven by using that term in our person-focused condition.

We did not see a moderating effect of confronter race. Past mixed findings (Czopp & Monteith, 2003; Czopp et al., 2006; Winslow, 2004) on the role of confronter race may speak to boundary conditions for the effects of this variable. For example, the extent to which the race of the confronter strengthens or weakens reactions to confrontations might depend on the confronted individual’s pre-existing attitudes toward that particular racial group or the frequency with which they interact with individuals who belong to that racial group. Further research will need to be conducted with relevant moderators to resolve these mixed findings.

We did find, however, differences in the race of the targeted group in the incident being confronted. Specifically, confrontations concerning anti-Black racism were reported as less person-focused than other types of racism and there was less of a benefit to the confronter when the confrontation was more behavior-focused. These exploratory findings suggest that when confronting anti-Black racism, the confronter will experience fewer benefits from using a behavior focus. That is, for other types of racism, a behavior focus is associated with both increased prosocial responding and decreased withdrawal/avoidant responding. For anti-Black racism, a behavior focus is also associated with increased prosocial responding (suggested by the main effect), but is not associated with decreased withdrawal/avoidant responding. Although causal links cannot be determined using a retrospective account methodology, these analyses suggest that confronting anti-Black racism may be more of a difficult task for the confronter, in terms of eliciting a less destructive response, compared to other types of racism.

It is important to note that our additional analyses using two other confrontation perspectives further replicate some findings but not others. Using retrospective accounts of confrontations from the confronter and witness perspectives, we see more evidence to suggest that a focus on the person results in more destructive types of responding. These analyses, however, did not support the link between behavior-focused confrontations and prosocial responding. From the witness perspective, it is possible that those who are not directly involved in the interaction missed some of the responses that occurred. Further, confronters may be less likely to remember incidents where they endured a constructive response to confronting someone, whereas confronted individuals may be more likely to remember the times that they reacted constructively (e.g., prosocially). Indeed, when comparing the means for these two samples (on a responses scale from 1 to 6), those confronted did report engaging in more prosocial behaviors ($M = 3.01, SD = 1.72$) than confronters reported receiving ($M = 2.51, SD = 1.57; t(357) = 2.89, p = .004$). Although we focus on the perspective that matches our first study (i.e., being confronted), the possibility of socially desirable responding is a potential limitation of Study 2. We address this by controlling for social desirability in a third study.

Both Study 1 and Study 2 did not specifically focus on the workplace, as Study 1 was a laboratory experiment and participants in Study 2 were allowed to answer questions about any confrontation incident, whether it occurred in the workplace or outside of it. In addition to addressing social desirability concerns, our third study focuses on participants' retrospective accounts of a time they were confronted while at work. In Study 3, we also address another limitation to Studies 1 and 2—the lack of explanatory mechanism. We have evidence to support that the focus of the confrontation is associated with different types of responses, but we do not

know why. In Study 3, we examine interpersonal cognitions and motivations that may explain the effects of confrontation focus.

Study 3

In Study 3 we employ the same retrospective account methodology as Study 2, but restrict race-based confrontation incidents to those that occurred in the workplace. We also wanted to address the Study 2 finding that individuals who were confronted, on average, reported responding in more prosocial ways to confrontations than confronters reported experiencing. This could mean that confronted individuals were engaging in socially desirable responding, and as such, we add a social desirability control to this study. We again examine the extent to which confrontation focus relates to different reactions to being confronted (H1 and H2). We also examine the extent to which race of the confronter impacts how people respond to confrontations (H5) and explore differences based on the race of the targeted group (RQ). Further, we propose and test two mediators of the relationship between confrontation focus and confrontation response: cognition and motivations.

Confrontations can be embarrassing and even hurtful, and confronted individuals can adopt very different types of cognitions and motivations in response. As stated earlier, Smart Richman and Leary (2009) theorize that reactions to experiences of relational devaluation are based, in part, on expectations of relationship repair. According to their model, when one believes they can repair a relationship they are more likely to act prosocially, whereas when one believes the relationship is not salvageable, they are more likely to act antisocially or to withdraw. As such, relationship repair efficacy may be one of the mechanisms explaining the relationship between confrontation focus and confrontation responses.

Research on reactions to transgressions suggests that some people respond to being harmed by others with benevolent (i.e., conciliatory, forgiving) motivations, whereas others respond with motivations to enact revenge (McCullough et al., 1998; McCullough & Hoyt, 2002). In addition to relationship repair beliefs, Bradfield and Aquino (1999) theorize relational motivations as mediators that explain why certain types of “personal offenses” at work lead to more constructive or destructive actions. Specifically, they theorize (and empirically support) that for someone who feels offended by an incident at work, severe offenses committed by unlikeable individuals result in more blame put on the offender and increase thoughts of revenge whereas milder offenses committed by likeable individuals result in less blame and increase thoughts of conciliation. In turn, these relational motivations result in subsequent revenge or forgiveness behavior.

Although our focus is not on confrontation severity or confronter likeability, we believe relational motivations will function similarly when comparing confrontation foci. That is, person-focused confrontations will be seen as more offensive than behavior-focused confrontations, and thus will be more likely to engender revenge motivations which will then predict antisocial and withdrawal behaviors. Conversely, behavior-focused confrontations will be seen as milder, leading to a greater likelihood of more benevolent relationship motivations followed by prosocial behaviors. Further bolstering these predictions, research on the benefits of forgiveness on personal well-being have also found evidence for reverse causality – people who feel good about themselves tend to have increased motivation for benevolence toward transgressors and decreased motivation for revenge (Bono et al., 2008; Orth et al., 2008). As previous work links person-focused feedback attributions to a sense of hopelessness (e.g., Hui et al., 2012; Joiner, 2001) and behavior-focused feedback to persistence (Foll et al., 2006; Ilgen &

Davis, 2000), it is likely that confronted individuals are in a better mental state to engage in benevolence and resist a desire for revenge in the latter type of confrontation situation, as opposed to the former.

Overall, we predict that when experiencing a behavior-focused confrontation, individuals will engage in more prosocial behavior because they both believe the relationship can be repaired and are motivated to repair it rather than harm it further. Conversely, we predict that when experiencing a person-focused confrontation, individuals will engage in more antisocial and withdrawal/avoidant behavior because they do not believe the relationship can be repaired and they are motivated to harm the relationship further rather than repair it. Specifically, we predict that:

Hypothesis 6: The relationship between behavior-focused confrontations and prosocial responding will be mediated by greater beliefs the relationship can be repaired (i.e., relationship repair efficacy), greater motivation to repair it (i.e., benevolent motivation), and lower motivation to harm the relationship further (i.e., revenge motivation)

Hypothesis 7: The relationship between person-focused confrontations and antisocial and withdrawal/avoidant responding will be mediated lower relationship repair efficacy, lower benevolent motivation, and greater revenge motivation

Participants and Procedure

We recruited participants via the online survey platform MTurk. We specifically advertised for individuals who had been confronted at work for saying/doing something racially insensitive and they had to be able to describe it in at least three sentences and in sufficient detail (e.g., a general summary of what happened and how they reacted). This requirement for payment allowed us to effectively screen out “bots” or individuals who did not fit the criteria.

An initial sample of 420 participants responded affirmatively to having experienced at least one of the types of confrontation experiences, provided an actual written response (filtering out those who said “none”, “N/A”, etc), and passed a thorough check to weed out automated “bot” responses (e.g., random series of letters, words, or numbers; copied text from the internet). This thorough check consisted of one undergraduate research assistant and one principal researcher reading every account and flag suspicious submissions. The principal researcher made the final call on all exclusions. Of those, 43 participants were excluded for providing incomplete information (e.g., saying only “I was confronted” with no details) and 135 participants were excluded for discussing an incident that did not fit the criteria (e.g., took place outside of work, described a time when the participant experienced racism, described another type of confrontation). After those exclusions and an additional exclusion of 12 participants who did not pass attention check items, we conducted analyses with a sample of 230 participants (54.76 % of original respondents). Although we excluded many potential participants, we used stringent criteria to ensure that the 230 participants included in analyses (97 female, mean age = 37.53; SD = 11.07) provided an actual incident of confrontation in order to increase our study’s validity.

Measures

Measures of confrontation focus (behavior- and person-focused), as well as confrontation response (prosocial, $\alpha = .91$; antisocial, $\alpha = .69$; withdrawal/avoidant, $\alpha = .77$) were identical to those used in Study 2.

Relationship Repair Efficacy

To measure relationship repair efficacy, we adapted Bradbury’s (1990) measure of relationship self-efficacy, which focused on one’s perceived ability to repair relationship conflict. We adapted the scale to be focused on a specific (rather than ongoing conflict). Items

began with “At the time of the confrontation, I felt” and included: I had little control over the reaction of this person, there was no way I could have solved their problem with me, I was helpless in dealing with this situation with this person, I felt I had no say over the conflict between us, I was able to do the things needed to settle this conflict (only non-reverse-coded item), There is little I could have done to resolve this conflict. All items save one were reverse-coded so that higher values indicate greater efficacy to repair the relationship. All six items ($\alpha = .85$) used a 1 (*Strongly Disagree*) to 6 (*Strongly Agree*) response scale.

Interpersonal Motivations

To measure relationship repair-related motivations following the confrontation, we adapted McCullough and colleagues’ (1998; 2002) Transgression-Related Interpersonal Motivations Inventory, focusing on benevolence motivation (example item: “At the time of the confrontation, I wanted us to have a positive relationship again”) and revenge motivation (example item: “At the time of the confrontation, I wanted to make this person pay”). We adapted items so that they were specifically framed as one’s motivations related to being confronted and chose not to administer one benevolence item (“I forgave this person”) that we thought would be confused as one’s actual behavior (offering forgiveness) rather than an internal state. Benevolence motivation ($\alpha = .86$) included six items and revenge motivation ($\alpha = .95$) included five. Both measures used a 1 (*Strongly Disagree*) to 6 (*Strongly Agree*) response scale.

Social Desirability

To control for socially desirable responding in our analyses, we included the Stöber’s (1999; 2001) Social Desirability Scale-17, which includes 17 true/ false items that assess socially desirable responding (example items: “In traffic I am always polite and considerate of others”; “I always eat a healthy diet”). This scale uses more updated language than other popular social

desirability scales (Crowne & Marlowe, 1960) and has shown satisfactory test-retest reliability, internal reliability, and convergent validity (Stöber, 1999) as well as discriminant validity (Stöber, 2001). We chose not to measure one item (“I have tried illegal drugs”) from the original scale due to findings from the scale’s creator (Stöber, 2001) that this item did not significantly correlate with the other items. The sixteen remaining items were scored (0 = non socially desirable answer, 1 = socially desirable answer) and summed to give each participant a social desirability score that could range from 0 to 16.

Additional Measures

Participants were asked several other questions about the confrontation, including these same questions from Study 2: their personal demographics (gender, race, age, typical working hours), how well they knew the confronter at the time of the confrontation, the confronter’s race/ethnicity, the race/ethnicity of the group toward which the confronter thought the participant was being racially insensitive, and whether or not the confronter used the term “racist” during the confrontation. Additionally, we asked participants about the work role of the person who confronted them (*coworker, supervisor/boss, subordinate/employee, client/customer, or other*).

Study 3 Results

Workplace Incident Characteristics

Characteristics of the participants and the incidents collected are available in Table 4. Of note, the majority (66.96 %) involved a coworker confronting the participant. Similar to Study 2, the most commonly reported race of the confronter was Black (45.22 %) and the most commonly targeted race was also Black (53.04 %). At the time of the confrontation, participants reported being employed in a variety of occupations, with the four most common being business and

financial operations (14.78 %), sales and related occupations (14.78 %), computer and mathematical occupations (10.43 %), and educational and library occupations (7.83 %).

Main Analyses

Table 7 provides the correlations of all relevant variables. To test for main effects (H1 and H2) as well as mediation (H6 and H7), three separate mediation models were conducted using the PROCESS Macro in SPSS (Preacher & Hayes, 2004). In the first model, prosocial responding was regressed on behavior-focused confrontation followed by three mediator variables (relationship efficacy, benevolence motivation, and revenge motivation). For the second and third models, antisocial and withdrawal/avoidant responding were separately regressed on person-focused confrontation followed by the same three mediator variables. In all three models, the other focus variable (e.g., behavior-focused confrontation for antisocial responding model), whether or not the term “racist” was used during the confrontation (0 = *no*, 1 = *yes*), and the Social Desirability scale score were entered as control variables. Power analyses indicated we exceeded the approximately 220 participants we need to detect mediation with parallel mediators given medium-sized correlations between variables and .80 power (Schoemann et al., 2017). As seen in Figure 2, the total effect of behavioral focus on prosocial responding was significant ($b = .19, p = .034$). Further the total effects of person focus on antisocial ($b = .26, p < .001$) and withdrawal/avoidant ($b = .21, p = .005$) responding were also significant, supporting H1 and H2.

The relationship between behavioral focus and prosocial responding was mediated by benevolent motivation (indirect effect = .13[.038, .229]), with behavioral focus associated with greater benevolent motivation ($b = .18, p = .007$) and benevolent motivation associated with

greater prosocial responding ($b = .69, p < .001$). The relationship was not mediated by relationship efficacy or revenge motivation. H6 was partially supported.

The relationships between person focus and antisocial and withdrawal/avoidant responding were both mediated by relationship repair efficacy (indirect effect = .03[.004, .065]; indirect effect = .05[.011, .117]; respectively) and revenge motivation (indirect effect = .14[.072, .208]; indirect effect = .09[.044, .148]; respectively). Neither relationship was mediated by benevolence motivation. Person focus was negatively associated with relationship repair efficacy ($b = -.15, p = .008$) and positively associated with revenge motivation ($b = .26, p < .001$). Relationship repair efficacy was negatively associated with both antisocial ($b = -.17, p = .005$) and withdrawal/avoidant ($b = -.36, p < .001$) responding. Revenge motivation was positively associated with both outcomes ($b = .50, p < .001$; $b = .33, p < .001$; respectively). Overall, H7 was partially supported. Similar to Study 2 findings, we found no significant interactions between confronter-target race similarity and confrontation focus, failing to support H5.

Exploratory Analyses.

Again, we found that the majority of our collected incidents (53.04 %) concerned confronting anti-Black racism. We compared anti-Black racism to all other types of racism on predictor, mediator, and outcome variables. Unlike Study 2, we did not find any differences in the extent to which the incidents were viewed as person-focused. We found two differences in the mediator variables. Specifically, anti-Black racist incidents were associated with lower reported relationship repair efficacy ($M = 2.85, SD = 1.19$), but also lower revenge motivation ($M = 2.00, SD = 1.32$) compared to reported relationship repair efficacy and revenge motivation for other types of racist incidents ($M = 3.25, SD = 1.17$; $M = 2.43, SD = 1.45$; respectively). There were no other differences in main study variables. We also explored if target of prejudice

(anti-Black racism versus other types of racism) interacted with confrontation focus variables to predict any of the three confrontation responses, as we did in Study 2. Targeted race of the incident did not interact with either confrontation focus variable to predict any of the three responses.

We also examined the relationship between person-focused confrontations and prosocial responding, as well as the relationship between behavior-focused confrontations and antisocial and withdrawal/avoidant responding. Person-focused confrontations were not associated with prosocial responding ($b = -.13, p = .083$) and there was no evidence of mediation for any of the cognition and motivation variables. We found that behavior-focused confrontations were associated with lower antisocial responding ($b = -.19, p = .012$) but not lower withdrawal/avoidant responding ($b = -.03, p = .729, ns$). We found a significant indirect effect of revenge motivation for both outcomes (indirect effect = $-.14[-.246, -.066]$; indirect effect = $-.09[-.177, -.040]$; respectively), with behavior-focused confrontations relating to lower revenge motivation ($b = -.28, p < .001$), which was associated with greater antisocial ($b = .50, p < .001$) and withdrawal/avoidant ($b = .33, p < .001$) responding. Further, we found a significant indirect effect of benevolent motivation for withdrawal/avoidant responding (indirect effect = $-.07[-.139, -.022]$), with behavior-focused confrontations relating to greater benevolent motivation ($b = .18, p = .007$), which was associated with lower withdrawal/avoidant responding ($b = -.38, p < .001$).

Study 3 Discussion

In a third study examining confrontation focus and confrontation response, we found additional support for the link between person focus and negative responses (i.e., antisocial and withdrawal/avoidant), as well as additional support for the link between behavioral focus and positive prosocial responses. This study was restricted to workplace incidents, suggesting that

these relationships between variables are found in and out of the workplace (Study 2 included many contexts) and outside of a controlled experiment environment (Study 1).

One of the main predictors of prosocial versus antisocial or withdrawal/avoidant responding posited by Smart Richman and Leary (2009) and Bradfield and Aquino (1999) was the expectation of being able to repair the relationship and relational motivations. Consistent with that framework, we found evidence to support that cognition (efficacy beliefs) and motivations (benevolence and revenge) related to relationship repair served as explanatory mechanisms for the focus – response relationship. Specifically, we found evidence that benevolent motivation explains the relationship between behavioral focus and prosocial responding. Further, revenge motivation and relationship repair efficacy explain the relationships between person focused confrontations and antisocial and withdrawal/avoidant responding. Taken together, these findings suggest that when we are confronted based on our behavior, we are motivated to repair the relationship with the individual (benevolent motivation) and subsequently engage in more prosocial responding. When we are confronted in a way that makes it feel like our character or person is being judged, we are less likely to believe the relationship can be repaired and more likely to be motivated to harm the relationship further (revenge motivation), predicting both subsequent antisocial and withdrawal responding.

Again, we found that confronter race did not moderate the focus-response relationship. We also did not replicate the Study 2 finding that anti-Black racism was less likely to result in reductions in withdrawal responses when there was greater perceived behavioral focus. We found some differences in cognitions and motivations when comparing anti-Black racism to other types of racism, with lower beliefs in the relationship being able to be repaired, but also less reported revenge motivation. Taken together with Study 2 findings, there are clearly some

nuances in confrontations across different types of racist transgressions that need to be further explored in order to get a fuller picture. However, overall findings also show more similarities than differences across different targeted racial groups, suggesting that confronting anti-Black racism largely functions similarly to confronting other types of racism.

An important limitation to note in this interpretation of Study 3 findings is that we employed a cross-sectional design, opening the possibility to alternative explanations of the relationships between variables. Our mediation findings could be a result of reverse causality, although motivations and cognitions as mechanisms to explain response to offenses/transgressions is theoretically rooted (e.g. Bradfield & Aquino, 1999; Smart Richman & Leary, 2009). Further, our main effect findings largely replicate Study 1 and 2 findings, providing triangulating evidence for those specific hypotheses.

General Discussion

Across three studies, we examined the extent to which person-focused confrontations elicited different responses than behavior-focused confrontations. Based on research and theory on interpersonal rejection motivations, causal attributions, and self-conscious emotions, we posited that behavior-focused confrontations would elicit more prosocial responses from those confronted, whereas person-focused confrontations would elicit more antisocial and withdrawal/avoidant responses. We examined this through an experimental study of anti-Black racism in which confrontation focus was manipulated (Study 1) and two survey studies in which confrontation focus was measured concerning real-life, retrospective accounts of confrontations related to racism more generally (Studies 2 and 3). We also were able to examine the race of the confronter and the race targeted in the incident as moderators (Studies 2 and 3) and test

relationship-related cognitions and motivations as mediators (Study 3). Across these studies, our findings highlight some key patterns that help us better understand race-based confrontations.

Theoretical Implications

When examining immediate reactions to confrontations through the use of a controlled laboratory paradigm (Study 1) and through analysis of real-life incidents (Studies 2 and 3), we see a consistent pattern of person-focused confrontations being associated with antisocial and withdrawal/avoidant responses. These findings contribute theoretically to the research on prejudiced-based confrontations. Although past research found that “extreme” or “hot” confrontations elicit more negative evaluations (Czopp et al., 2006; Shultz & Maddox, 2013), our studies focused on interpersonal behavioral responses and were grounded in the causal attribution and interpersonal rejection literatures. As such, we provide some needed specificity as to what types of confrontations might be more or less effective in terms of specific types of responses.

Although we did not find differences in reported emotions after a more person-focused confrontation (Study 1), we did find differences in interpersonal cognitions and motivations, which mediated the relationship between confrontation focus and confrontation responses. Specifically, the extent to which one felt the relationship could be fixed (relationship efficacy) and the extent to which one was motivated to further hurt the relationship (revenge motivation) explained the relationship between person-focused confrontations and antisocial and withdrawal/avoidant responding. These findings are in line with Smart Richman and Leary’s (2009) interpersonal rejection model in that confrontations that convey limited opportunity to repair the situation or relationship because of fault with *one’s being* are associated with less constructive responses. We also find a good amount of support for the link between behavior-

focused confrontations and prosocial responses in all three studies, with the caveat that these findings were not consistent across all three perspectives in Study 2 (see Study 2 Discussion). These findings provide tentative support for the relatively positive effects of behavior-focused confrontations, as participants who received this type of confrontation in the lab (Study 1) and who reported about prior receipt of this type of confrontation (Studies 2 and 3) both responded in a more prosocial fashion. Further supporting the interpersonal rejection framework, the relationship between a more behavior-focused confrontation and prosocial responding was explained by an increase in motivation toward repairing the relationship (i.e., benevolence motivation, Study 3). Interestingly, experiencing a more behavior-focused confrontation was not associated with greater beliefs that the relationship could be saved, but participants confronted in this way were more motivated to try to save it.

Emotions were not differentiated by person versus behavior conditions in Study 1. One possibility to explain these findings is psychometric limitations of the emotions measures we used in Study 1. However, we did find that both types of confrontations elicited greater guilt and shame than the control no-confrontation condition, suggesting that the measure we used was able to capture differences in emotions. Alternatively, our findings may indicate that emotions are not as key to understanding interpersonal responses to confrontations as are one's post-confrontation beliefs and motivations regarding the interpersonal relationship.

We failed to find support for the role of confronter race as an amplifier of the confrontation focus – response link in either Studies 2 or 3. These null findings suggest that any effects of confronter race might be more complicated than simply a more heightened response to confrontations from target group members. Although past work has found that individuals feel worse about themselves when confronted by a target group member (Czopp et al., 2006) and are

more worried about appearing racist to a target group member (Winslow, 2004), there is other evidence to the contrary (Czopp & Monteith, 2003). We do not posit that confronter race is irrelevant to confrontation responses, but suggest that there are other unmeasured factors that might mask effects. For example, the perceived credibility of a messenger can determine their influence on others' attitudes (Tormala et al., 2007). This could suggest that minority confronters will be less influential to majority members, as outgroup members are generally viewed as less credible than ingroup members (Clark & Maass, 1988), but it is also possible that in this specific context, target group members are seen as having more expertise in regards to racism toward their own group. There is also evidence to suggest that for some individuals, minority group messengers will have more of an influence on attitudes (regardless of credibility), particularly when they are personally motivated to avoid being prejudiced (i.e., the "watchdog hypothesis", Petty et al., 1999). Status and power may also be relevant. Even when the confronter is from a minority group associated with lower societal power (and thus might be thought to be pervasively less influential), that does not mean they do not have alternative sources of power within that specific context that are unrelated to their marginalized identity (e.g., having a high status position; Lyons et al., 2017). Unfortunately, we do not have a large enough sample size to explore interactions between the rank of the confronter (i.e., coworker, boss) and confronter race. Future research should examine the perceived credibility, status, and power of the confronter, along with other characteristics of the confronter and confronted, as boundary conditions for the role of confronter race in confrontation responses.

Exploratory findings in Studies 2 and 3 suggests some potential differences in confronting anti-Black racism compared to other types of racism. Specifically, Study 2 findings suggest that withdrawal/avoidant responding might depend on what type of racism is being

confronted. For all types of racism, reduced withdrawal/avoidant responding was associated with less person-focused confrontations (i.e., the main effect finding). However, a behavior-focused confrontation was only associated with reduced withdrawal/avoidant responding for racism that targeted non-Black individuals (i.e., the interactive effect, see Figure 1). In essence, Study 2 results suggest that it may be particularly difficult to reduce negative reactions to confronting anti-Black racism. Why might this be? If individuals tend to view anti-Black racist behavior as a more serious transgression or more indicative of internal racist beliefs than other types of racist behavior, then that might explain why even a behavioral focus was not associated with less withdrawal /avoidant responding. There is some research to suggest that the prototypical racist interaction is that of a White perpetrator and a Black victim (Inman & Baron, 1996) and that one's prototypes of racism inform reactions to ambiguous incidents (Corning & Bucchianeri, 2010; Rodin et al., 1990), although this work did not include scenarios with victims from other minority races/ethnicities. If anti-Black racism is seen as most prototypical of a racist act, confrontations involving this type of racism may often be particularly threatening to one's sense of self and, as such, less effective in reducing withdrawal/avoidant responding.

Importantly, this finding was not replicated in Study 3, although there were some differences found for confronting anti-Black racism compared to other types of racism. Namely, confronting anti-Black racism was associated with both lower beliefs that the relationship could be repaired (i.e., relationship efficacy) but also lower revenge motivation. These complex findings necessitate future research that can explore the boundary conditions and nuances associated with confronting anti-Black racism. However, based on our Study 1 experiment that only focused on anti-Black racism, and our main and mediation effects findings in Studies 2 and

3 that held for all targeted racial groups, our findings provide insights into the role of focus, relationship beliefs, and relationship motivations in confronting anti-Black racism.

Our findings suggest that a deeper integration of the race-based confrontations research and work on interpersonal rejection would be a fruitful avenue for future research. Specifically, our findings suggest that confrontation focus will affect one's perceived ability to repair a relationship and motivation to do so, which will lead to more/ less constructive responses. Smart Richman and Leary's (2009) model discusses several other potential predictors of reactions to rejection which could be examined in the context of race-based confrontations. For example, they posit that a highly valued relationship with the rejecter will lead to more prosocial responses and that chronic rejection leads to withdrawal/avoidant. In the context of confrontations, this would imply that close others will receive more prosocial responses when confronting and that individuals who continue to confront will end up being avoided. These are all important questions that could be addressed in future research on race-based confrontations.

Practical Implications

Practically, findings from this research can be used to inform diversity practices in workplaces. First, people are often hesitant to confront prejudice in the workplace in real time, even if they think they would (Woodzicka & LaFrance, 2001), for fear of backlash (Shultz & Maddox, 2013) and out of an aversion to conflict (Stevens & Fiske, 1995). Ashburn-Nardo and colleagues (2008) point out that particularly for those with majority social identities, many people do not have experience thinking about race or knowing how to confront it, and might even think it is a lost cause to do so. Applying their Confronting Prejudiced Responses (CPR) model to organizations, they suggest that employees need to engage in exercises where they can model confronting, so they are more likely to do so in the future. Our findings suggest that

behavioral modeling exercises should avoid inclusion of confrontations that evoke global, stable, and personal attributions emphasizing a person's problematic character, and instead focus more on confrontations that evoke specific, temporary, and external attributions to behavior, if the goal is to reduce backlash to confrontations.

Although changing the ways in which employees confront prejudice could be helpful in increasing inclusion and curtailing future bad behavior, senior leaders in organizations also play an important role in creating a safe environment for BIPOC employees through compassionate acknowledgement of racially traumatic events (McCluney et al., 2017). Our results suggest that organizational leaders need to work on reducing negative responses in confronted employees, as we did see a good amount of antisocial and withdrawal/avoidant behavior in participants' immediate responses. Although it is typical and understandable to feel bad when confronted about one's own prejudices, being open to criticism and willing to learn about one's own gaps in understanding can help create an organizational climate in which employees can have productive conversations about prejudice and bias. Our Study 3 findings in particular suggest that organizational leaders could provide concrete and actionable advice to a confronted individual on how to make amends, as a way to increase relationship repair beliefs and thus potentially reduce destructive responding. Organizational leaders could also instruct confronted individuals to take some time to think about any positive aspects of the confrontation (e.g., opportunity for personal growth), as doing so has been shown to increase benevolent motivation and decrease revenge motivation when an individual feels harmed (McCullough et al., 2006). Strategies to increase expectancy and desire for relationship repair could perhaps be integrated into formal organizational procedures for responding to bias in the workplace. Overall, our research findings

point to the importance of considering confrontation messaging and responses to confrontations when attempting to create an inclusive climate.

Potential Limitations

As in all research, there are potential limitations to our findings. Our controlled lab paradigm improves the internal validity of our inferences, but our participants and the confederate were strangers and thus had less of an incentive to work together effectively, which differs from many real-world contexts. Research on confrontations is still in its nascent stages, warranting a controlled environment for studying the effects of key variables on interactions. The paradigm we used permitted the examination of confrontation responses while eliminating confounding and extraneous variables (e.g., the type of prejudiced comment, the power of the person confronting). Further, in Studies 2 and 3 we examined confrontation responses using a methodology with greater external validity (actual retrospective accounts of confrontations) and found similar main effects.

Studies 2 and 3 could be limited by threats to internal validity, such as social desirable responding and memory distortions. Study 3 controlled for social desirability and found similar main effects to the other studies. Further, in looking at Study 3 data, many individuals reported their engagement in undesirable, antisocial behavior. Approximately 40% of Study 3 participants agreed to some degree that they “became angry”, ~50% reported engaging in denial, and ~20% reported yelling or becoming aggressive. Still, with a retrospective design, there is always a possibility of memory bias and the potential for a reverse causality explanation to findings. We reduced the likelihood of this happening asking participants to describe, in detail, an incident in which they were confronted before knowing which items they would be responding to about said incident. This research design makes it more difficult to recraft the narrative once they are aware

of the questions they will be asked about the incident. Yet, it is possible that individuals actually encoded and remembered an event in a way that makes them feel better about themselves. Based on the distastefulness of antisocial behavior, the finding that might be most affected by this potential limitation would be the antisocial findings (i.e., I remember the person as attacking my character to make me feel better about reacting poorly). However, we do not think this is a likely explanation for our findings, as the relationship between person focus and antisocial responding was also found when surveying confronters and witnesses to confrontations. Taking the three studies together, we have some triangulating evidence that links confrontation focus to prosocial, antisocial, and withdrawal/avoidant responses in both experimental and survey data, as well as with student and working adult populations.

Another potential limitation is our focus on immediate reactions of the confronted. We see the importance of understanding initial reactions to being confronted, as researchers and activists alike have put forth confronting as an important tool for combatting racism (e.g., Ashburn-Nardo et al., 2008; Saad, 2020), yet in-the-moment destructive reactions will likely inhibit many from attempting to confront others in the future. However, our results cannot speak to other important outcomes relevant to confronting prejudice, such as how much the confronted individual has actually engaged in self-reflection, how other parties perceive confrontations, or whether the confronted individual expresses less prejudice over time. For example, Martinez and colleagues (2017) focused on confronting heterosexism and found that witnesses reacted negatively to confronters being more direct (similar to person-based) when their tone was hostile, but not when their tone was calm. Future research exploring how race-based confrontations affect the confronter, bystanders, and the organization as a whole will give us a more complete picture of the confrontation process.

Conclusion

Making individuals aware of their own prejudices can be a first step in reducing biases and prejudiced behavior, but confronting someone about racism can result in antisocial or avoidant responses. Across three studies, we found evidence that the focus of confrontations (behavior vs. person) matters in terms of how recipients immediately react toward the confronter. We also found that relationship-related cognitions and motivations help explain these findings, as well as some differences in confronting anti-Black racism compared to confronting other types of racism. Overall, our findings suggest that deconstructive responses are not a certainty, and that the focus of the confrontation can make constructive responses possible.

Footnote

¹The two conditions from Czopp and colleagues (2006) were worded as follows: low threat/ “cold”: *but maybe it would be good to think about Blacks in other ways that are a little more fair? it just seems that a lot of times Blacks don't get equal treatment in our society. you know what i mean?*; high threat/ “hot”: *but you should really try to think about Blacks in other ways that are less prejudiced. it just seems that you sound like some kind of racist to me. you know what i mean?*

References

- Abramson, L.Y., Seligman, M. E., & Teasdale, J. D. (1978). Learned helplessness in humans: critique and reformulation. *Journal of Abnormal Psychology, 87*(1), 49-74. <https://doi.org/10.1037/0021-843X.87.1.49>
- Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A dynamic model of guilt: Implications for motivation and self-regulation in the context of prejudice. *Psychological Science, 18*(6), 524-530. <https://doi.org/10.1111/j.1467-9280.2007.01933.x>
- Anderson, C. A. (1983). The causal structure of situations: The generation of plausible causal attributions as a function of type of event situation. *Journal of Experimental Social Psychology, 19*(2), 185-203. [https://doi.org/10.1016/0022-1031\(83\)90037-9](https://doi.org/10.1016/0022-1031(83)90037-9)
- Anderson, C. A., Miller, R. S., Riger, A. L., Dill, J. C., & Sedikides, C. (1994). Behavioral and characterological attributional styles as predictors of depression and loneliness: Review, refinement, and test. *Journal of Personality and Social Psychology, 66*(3), 549-558. <https://doi.org/10.1037/0022-3514.66.3.549>
- Andrews, G. R., & Debus, R. L. (1978). Persistence and the causal perception of failure: Modifying cognitive attributions. *Journal of educational psychology, 70*(2), 154. <https://doi.org/10.1037/0022-0663.70.2.154>
- Ashburn-Nardo, L., Morris, K. A., & Goodwin, S. A. (2008). The confronting prejudiced responses (CPR) model: Applying CPR in organizations. *Academy of Management Learning & Education, 7*(3), 332-342. <https://doi.org/10.5465/amle.2008.34251671>

- Bono, G., McCullough, M. E., & Root, L. M. (2008). Forgiveness, feeling connected to others, and well-being: Two longitudinal studies. *Personality and Social Psychology Bulletin*, 34(2), 182-195.
<https://doi.org/10.1177/0146167207310025>
- Bradfield, M., & Aquino, K. (1999). The effects of blame attributions and offender likableness on forgiveness and revenge in the workplace. *Journal of Management*, 25(5), 607-631.
<https://doi.org/10.1177/014920639902500501>
- Brambilla, M., & Leach, C. W. (2014). On the importance of being moral: The distinctive role of morality in social judgment. *Social Cognition*, 32(4), 397-408.
<https://doi.org/10.1521/soco.2014.32.4.397>
- Clark III, R. D., & Maass, A. (1988). The role of social categorization and perceived source credibility in minority influence. *European journal of social psychology*, 18(5), 381-394.
<https://doi.org/10.1002/ejsp.2420180502>
- Clark, R., Anderson, N. B., Clark, V. R., & Williams, D. R. (1999). Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist*, 54(10), 805-816.
<https://doi.org/10.1037/0003-066X.54.10.805>
- Corning, A. F., & Bucchianeri, M. M. (2010). Perceiving racism in ambiguous situations: Who relies on easy-to-use information?. *The Journal of Social Psychology*, 150(3), 258-277.
<https://doi.org/10.1080/00224540903365398>
- 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 confrontation. *Journal of Personality and Social Psychology*, 90(5), 784-803.
<https://doi.org/10.1037/0022-3514.90.5.784>
- 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), 424-435.
<https://doi.org/10.1037/0022-3514.36.4.424>
- Esposito, S. R., Hornsey, M. J., & Spoor, J. R. (2013). Shooting the messenger: Outsiders critical of your group are rejected regardless of argument quality. *British Journal of Social Psychology*, 52(2), 386-395. <https://doi.org/10.1111/bjso.12024>
- Federal Bureau of Investigation. (2020). *Hate Crimes Statistics, 2019: Victims*. Retrieved from <https://ucr.fbi.gov/hate-crime/2019/topic-pages/victims.pdf>.
- Fiske, S. T., Cuddy, A. J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878-902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Foll, D. L., Rasche, O., & Higgins, N. C. (2006). Persistence in a Putting Task During Perceived Failure: Influence of State-attributions and Attributional Style. *Applied Psychology*, 55(4), 586-605.
<https://doi.org/10.1111/j.1464-0597.2006.00249.x>
- Fresco, D. M., Alloy, L. B., & Reilly-Harrington, N. (2006). Association of attributional style for negative and positive events and the occurrence of life events with depression and anxiety. *Journal of Social and Clinical Psychology*, 25(10), 1140-1160.
<https://doi.org/10.1521/jscp.2006.25.10.1140>
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148-168.

<https://doi.org/10.1037/a0034726>

- Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review*, 3(1), 5-39. <https://doi.org/10.1007/BF01323661>
- Hareli, S., & Hess, U. (2008). The role of causal attribution in hurt feelings and related social emotions elicited in reaction to other's feedback about failure. *Cognition and Emotion*, 22(5), 862-880. <https://doi.org/10.1080/02699930701541641>
- Heesacker, M., Petty, R. E., & Cacioppo, J. T. (1983). Field dependence and attitude change: Source credibility can alter persuasion by affecting message-relevant thinking. *Journal of Personality*, 51(4), 653-666. <https://doi.org/10.1111/j.1467-6494.1983.tb00872.x>
- Houston, D. M. (2016). Revisiting the relationship between attributional style and academic performance. *Journal of Applied Social Psychology*, 46(3), 192-200. <https://doi.org/10.1111/jasp.12356>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*.
- Hu, T., Zhang, D., & Yang, Z. (2015). The relationship between attributional style for negative outcomes and depression: A meta-analysis. *Journal of Social and Clinical Psychology*, 34(4), 304-321. <https://doi.org/10.1521/jscp.2015.34.4.304>
- Hui, C. H., Pak, S. T., Kwan, S. O., & Chao, A. (2012). Attributional style and engagement/disengagement responses in the Chinese workforce. *Applied Psychology*, 61(2), 204-226. <https://doi.org/10.1111/j.1464-0597.2011.00463.x>
- Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology*, 49(3), 550-565. <https://doi.org/10.1111/1464-0597.00031>

- Inman, M. L., & Baron, R. S. (1996). Influence of prototypes on perceptions of prejudice. *Journal of Personality and Social Psychology*, 70(4), 727-739. <https://doi.org/10.1037/0022-3514.70.4.727>
- Joiner Jr, T. E. (2001). Negative attributional style, hopelessness depression and endogenous depression. *Behaviour Research and Therapy*, 39(2), 139-149. [https://doi.org/10.1016/S0005-7967\(99\)00160-6](https://doi.org/10.1016/S0005-7967(99)00160-6)
- Kendi, I. X. (2019). *How to be an antiracist*. One world; New York, New York.
- Ledbetter, A. M. (2009). Measuring online communication attitude: Instrument development and validation. *Communication Monographs*, 76(4), 463-486. <https://doi.org/10.1080/03637750903300262>
- Liden, R. C., Ferris, G. R., & Dienesch, R. M. (1988). The influence of causal feedback on subordinate reactions and behavior. *Group & Organization Studies*, 13(3), 348-373. <https://doi.org/10.1177/105960118801300307>
- Lyons, B. J., Pek, S., & Wessel, J. L. (2017). Toward a "Sunlit Path": Stigma Identity Management As a Source of Localized Social Change Through Interaction. *Academy of Management Review*, 42(4), 618-636. <https://doi.org/10.5465/amr.2015.0189>
- Martinez, L. R., Hebl, M. R., Smith, N. A., & Sabat, I. E. (2017). Standing up and speaking out against prejudice toward gay men in the workplace. *Journal of Vocational Behavior*, 103, 71-85. <https://doi.org/10.1016/j.jvb.2017.08.001>
- McCluney, C. L., Bryant, C. M., King, D. D., & Ali, A. A. (2017). Calling in Black: a dynamic model of racially traumatic events, resourcing, and safety. *Equality, Diversity and Inclusion: An International Journal*, 36(8), 767-786. <https://doi.org/10.1108/EDI-01-2017-0012>
- McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology*

Bulletin, 28(11), 1556-1573. <https://doi.org/10.1177/014616702237583>

- McCullough, M. E., Rachal, K. C., Sandage, S. J., Worthington Jr, E. L., Brown, S. W., & Hight, T. L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology*, 75(6), 1586-1603. <https://doi.org/10.1037/0022-3514.75.6.1586>
- McCullough, M. E., Root, L. M., & Cohen, A. D. (2006). Writing about the benefits of an interpersonal transgression facilitates forgiveness. *Journal of Consulting and Clinical Psychology*, 74(5), 887-897 <https://doi.org/10.1037/0022-006X.74.5.887>
- McFarland, C., & Ross, M. (1982). Impact of causal attributions on affective reactions to success and failure. *Journal of Personality and Social Psychology*, 43(5), 937. <https://doi.org/10.1037/0022-3514.43.5.937>
- McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of personality and social psychology*, 53(2), 247-256. <https://doi.org/10.1037/0022-3514.53.2.247>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30-46. <https://doi.org/10.1037/1082-989X.1.1.30>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological methods*, 17(3), 437-455. <https://doi.org/10.1037/a0028085>
- Metalsky, G. I., Abramson, L. Y., Seligman, M. E., Semmel, A., & Peterson, C. (1982). Attributional styles and life events in the classroom: Vulnerability and invulnerability to depressive mood reactions. *Journal of Personality and Social Psychology*, 43(3), 612-617. <https://doi.org/10.1037/0022-3514.43.3.612>

- Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations*, 13(2), 183-200.
<https://doi.org/10.1177/1368430209353633>
- Nauta, M. M., Epperson, D. L., & Waggoner, K. M. (1999). Perceived causes of success and failure: Are women's attributions related to persistence in engineering majors?. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 36(6), 663-676. [https://doi.org/10.1002/\(SICI\)1098-2736\(199908\)36:6<663::AID-TEA5>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1098-2736(199908)36:6<663::AID-TEA5>3.0.CO;2-F)
- Oluo, I. (2019). *So you want to talk about race*. Hachette; UK.
- Orth, U., Berking, M., Walker, N., Meier, L. L., & Znoj, H. (2008). Forgiveness and psychological adjustment following interpersonal transgressions: A longitudinal analysis. *Journal of Research in Personality*, 42(2), 365-385. <https://doi.org/10.1016/j.jrp.2007.07.003>
- Peterson, C., & Barrett, L. C. (1987). Explanatory style and academic performance among university freshman. *Journal of Personality and Social Psychology*, 53(3), 603-607.
<https://doi.org/10.1037/0022-3514.53.3.603>
- Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The attributional style questionnaire. *Cognitive Therapy and Research*, 6(3), 287-299.
<https://doi.org/10.1007/BF01173577>
- Petty, R. E., Fleming, M. A., & White, P. H. (1999). Stigmatized sources and persuasion: Prejudice as a determinant of argument scrutiny. *Journal of Personality and Social Psychology*, 76(1), 19-34.
<https://doi.org/10.1037/0022-3514.76.1.19>
- Pitcan, M., Park-Taylor, J., & Hayslett, J. (2018). Black men and racial microaggressions at work. *The Career Development Quarterly*, 66(4), 300-314. <https://doi.org/10.1002/cdq.12152>

- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4), 717-731. <https://doi.org/10.3758/BF03206553>
- Reskin, B. (2012). The race discrimination system. *Annual Review of Sociology*, 38, 17-35. <https://doi.org/10.1146/annurev-soc-071811-145508>
- Rodin, M. J., Price, J. M., Bryson, J. B., & Sanchez, F. J. (1990). Asymmetry in prejudice attribution. *Journal of Experimental Social Psychology*, 26(6), 481-504. [https://doi.org/10.1016/0022-1031\(90\)90052-N](https://doi.org/10.1016/0022-1031(90)90052-N)
- Ruggs, E. N., Hebl, M. R., Law, C., Cox, C. B., Roehling, M. V., & Wiener, R. L. (2013). Gone fishing: I-O psychologists' missed opportunities to understand marginalized employees' experiences with discrimination. *Industrial and Organizational Psychology*, 6(1), 39-60. <https://doi.org/10.1111/iops.12007>
- Saad, L. F. (2020). *Me and White Supremacy: Combat Racism, Change the World, and Become a Good Ancestor.* Sourcebooks, Inc.
- Sanjuán, P., Pérez, A., Rueda, B., & Ruiz, Á. (2008). Interactive effects of attributional styles for positive and negative events on psychological distress. *Personality and Individual Differences*, 45(2), 187-190. <https://doi.org/10.1016/j.paid.2008.03.004>
- Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, 8, 379-386. <https://doi.org/10.1177/1948550617715068>
- 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>

- Smart Richman, L., & Leary, M. R. (2009). Reactions to discrimination, stigmatization, ostracism, and other forms of interpersonal rejection: a multimotive model. *Psychological review*, 116(2), 365.
<https://doi.org/10.1037/a0015250>
- Sommers, S. R., & Norton, M. I. (2006). Lay theories about White racists: What constitutes racism (and what doesn't). *Group Processes & Intergroup Relations*, 9(1), 117-138.
<https://doi.org/10.1177/1368430206059881>
- Steelman, L. A., & Rutkowski, K. A. (2004). Moderators of employee reactions to negative feedback. *Journal of Managerial Psychology*, 19(1), 6-18. <https://doi.org/10.1108/02683940410520637>
- Stevens, L. E., & Fiske, S. T. (1995). Motivation and cognition in social life: A social survival perspective. *Social Cognition*, 13(3), 189-214. <https://doi.org/10.1521/soco.1995.13.3.189>
- Sue, D. W., Capodilupo, C. M., & Holder, A. (2008). Racial microaggressions in the life experience of Black Americans. *Professional Psychology: Research and Practice*, 39(3), 329-336.
<https://doi.org/10.1037/0735-7028.39.3.329>
- Tangney, J. P., Miller, R. S., Flicker, L., & Barlow, D. H. (1996). Are shame, guilt, and embarrassment distinct emotions? *Journal of Personality and Social Psychology*, 70(6), 1256.
<https://doi.org/10.1037/0022-3514.70.6.1256>
- Tangney, J. P., Wagner, P., Fletcher, C., & Gramzow, R. (1992). Shamed into anger? The relation of shame and guilt to anger and self-reported aggression. *Journal of Personality and Social Psychology*, 62(4), 669. <https://doi.org/10.1037/0022-3514.62.4.669>
- Tobin, S. J., & Raymundo, M. M. (2009). Persuasion by causal arguments: The motivating role of perceived causal expertise. *Social Cognition*, 27(1), 105-127.
<https://doi.org/10.1521/soco.2009.27.1.105>

- Tormala, Z. L., Briñol, P., & Petty, R. E. (2007). Multiple roles for source credibility under high elaboration: It's all in the timing. *Social Cognition*, 25(4), 536-552.
<https://doi.org/10.1521/soco.2007.25.4.536>
- Wallace, L. E., Wegener, D. T., & Petty, R. E. (2020). When sources honestly provide their biased opinion: Bias as a distinct source perception with independent effects on credibility and persuasion. *Personality and Social Psychology Bulletin*, 46(3), 439-453.
<https://doi.org/10.1177/0146167219858654>
- Waung, M., & Jones, D. R. (2005). The Effect of Feedback Packaging on Ratee Reactions. *Journal of Applied Social Psychology*, 35(8), 1630-1652. <https://doi.org/10.1111/j.1559-1816.2005.tb02188.x>
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92(4), 548. <https://doi.org/10.1037/0033-295X.92.4.548>
- Weiner, B. (1986). *An attributional theory of motivation and emotion*, New York: Springer.
<https://doi.org/10.1007/978-1-4612-4948-1>
- Wenzel, M., Woodyatt, L., & McLean, B. (2020). The effects of moral/social identity threats and affirmations on psychological defensiveness following wrongdoing. *British Journal of Social Psychology*, 59(4), 1062-1081. <https://doi.org/10.1111/bjso.12378>
- Winslow, M. P. (2004). Reactions to the imputation of prejudice. *Basic and Applied Social Psychology*, 26, 289-297. https://doi.org/10.1207/s15324834basp2604_5
- 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>
- Zou, L. X., & Cheryan, S. (2017). Two axes of subordination: A new model of racial position. *Journal of Personality and Social Psychology*, 112(5), 696–717. <https://doi.org/10.1037/pspa0000080>

Table 1

Pilot Data Means and Standard Deviations

	Person-Focused Condition	Behavior-Focused Condition
Variable	Mean(sd)	Mean(sd)
External vs. Internal	6.39 ^a (1.01)	2.61 ^c (1.50)
Temporary vs. Stable	6.39 ^a (1.05)	2.69 ^c (1.30)
Specific vs. Global	6.19 ^a (1.18)	2.28 ^c (1.43)
	High Threat Condition	Low Threat Condition
External vs. Internal	4.63 ^b (1.60)	3.23 ^d (1.48)
Temporary vs. Stable	4.70 ^b (1.43)	3.53 ^d (1.34)
Specific vs. Global	4.35 ^b (1.74)	2.99 ^d (1.49)

Note. Lowest $N = 76$. Any value with a different superscript indicates difference from all other conditions. Higher values indicate the condition was judged to be more internal/ stable/ global.

Table 2

Correlations for Study 1 Variables of Interest

Variable	1	2	3	4	5	6	7	8	9	10	11		
Acknowledge (1)													
Bias Recognized (2)	.83												
Self-Improvement (3)	.59	.62											
Apology (4)	.22	.13	.34										
Appreciation/Thankfulness (5)	.50	.57	.64	.15									
Denial (6)	-.67	-.74	-.71	-.22	-.73								
Hostile (7)	-.42	-.41	-.39	-.45	-.48	.53							
Want to work with again (8)	.00	.11	.30	.20	.20	-.07	-.27						
Shame (9)	.24	.33	.19	.00	.08	-.13	.04	-.43					
Guilt (10)	.40	.50	.38	-.10	.24	-.30	-.30	-.30	.84				
Number Negative (9)	.23	.10	.09	.07	-.09	-.16	.08	-.21	.25	.21			
Participant Race (10)	.12	.14	.04	-.05	.20	.10	-.08	.32	-.25	-.16	-.04		
Participant Age (11)	-.10	-.05	-.12	-.25	.02	-.07	-.14	-.10	-.08	-.08	-.16	.09	
Participant Gender (12)	-.17	-.10	.02	-.16	-.13	.20	.14	-.01	.06	.10	-.05	.14	.14

Note. Lowest $N = 46$, for variables that were not measured in the control condition. Bolded values are significant at the $p < .05$ level. Number negative = number of stereotypic responses given. Participant Gender coded as 0 = female, 1 = male. Participant Race coded as 0 = White, 1 = non-Black POC.

Table 3

Unadjusted Means and Standard Deviations by Condition, Study 1

Outcome Variable	Behavior-Focused	Person-Focused	t-value	p-value
	(n = 22)	(n= 24)		
	Mean (sd)	Mean (sd)		
Prosocial Responses				
(overall mean)	3.44(1.23) ^a	2.42(1.08) ^b	-3.16	.003
Acknowledge	3.65(1.43) ^a	2.64(1.42) ^b	-2.79	.008
Bias Recognized	3.34(1.56) ^a	2.04(1.31) ^b	-3.24	.002
Self-Improvement	3.22(1.79) ^a	1.98(1.40) ^b	-2.68	.010
Apology	3.17(1.63)	3.15(1.81)	.03	.978
Appreciation/Thankfulness	3.84(1.82) ^a	2.27(1.67) ^b	-3.11	.003
Antisocial Responses				
(overall mean)	2.05(.89) ^a	3.35(1.53) ^b	3.64	.001
Denial	2.90(1.62) ^a	4.43(1.66) ^b	3.49	.001
Hostile	1.21(.26) ^a	2.27(1.83) ^b	2.69	.010

Note. Different superscripts indicate significant difference between conditions at the $p < .05$ level.

Table 4

Participant and Confrontation Characteristics, Studies 2 and 3

	Study 2 N (%)	Study 3 N (%)		Study 2 N (%)	Study 3 N (%)
Participant Gender			Confronter Race		
Female	97 (57.4)	103 (42.6)	Afr. Am./ Black	74 (43.8)	109 (45.0)
Male	72 (42.6)	137 (56.6)	East Asian	7 (4.1)	16 (6.6)
Other		2 (.8)	Hispanic/Latinx	22 (13.0)	17 (7.0)
Participant Race			Middle-Eastern	3 (1.8)	6 (2.5)
Afr. Am./ Black	20 (11.8)	34 (14.0)	NA/ AN	4 (2.4)	9 (3.7)
East Asian	9 (5.3)	10 (4.1)	Non-Hispanic White	47 (27.8)	68 (28.1)
Hispanic/Latinx	9 (5.3)	15 (6.2)	South Asian	1 (.6)	5 (2.1)
Middle-Eastern	1 (.6)	1 (.4)	Bi/Multi-Racial	2 (1.2)	9 (3.7)
NA/ AN	3 (1.8)	2 (.8)	Other	1 (.6)	0
Non-Hispanic White	114 (67.5)	161 (66.5)	Don't Know	7 (4.1)	3 (1.2)
South Asian	4 (2.4)	6 (2.5)	Targeted Race		
Bi/Multi-Racial	6 (3.6)	12 (5.0)	Afr. Am./ Black	96 (56.3)	129 (53.3)
Other	3 (1.8)	1 (.4)	East Asian	5 (3.0)	19 (7.9)
Participant Emp. Status			Hispanic/Latinx	21 (12.6)	23 (9.5)
Part-time (<30 h/w)	25 (14.8)	19 (8.3)	Middle-Eastern	1 (.6)	9 (3.7)
Full-time (>30 h/w)	80 (47.3)	210 (91.3)	NA/ AN	6 (3.6)	7 (2.9)
Retired	33 (19.5)	0	Non-Hispanic White	20 (12.0)	28 (11.6)
Other	31 (18.3)	1 (.4)	South Asian	3 (1.8)	5 (2.1)
Participant Knew Confronter			Bi/Multi-Racial	7 (4.2)	15 (6.2)
Not well at all	53 (31.4)	62 (25.6)	Other	3 (1.8)	2 (.8)
Somewhat well	42 (25.1)	80 (33.1)	Don't Know	5 (3.0)	5 (2.1)
Moderately well	29 (17.4)	65 (26.9)	Confronter and Target:		
Very well	45 (26.6)	35 (14.5)	Same Race	94 (56.6)	166 (69.5)
Confronter Work Role			Different Race	72 (43.4)	73 (30.5)
Participant's coworker	not asked	162 (66.9)	Term "racist" used?		
Participant's boss	not asked	21 (8.7)	Yes	105 (62.1)	169 (69.8)
Participant's customer	not asked	32 (13.2)	No	64 (37.9)	71 (29.3)
Participant's employee	not asked	23 (9.5)			
Other	not asked	4 (1.7)			

Note. Afr. Am. = African American, NA/AN = Native American/ Alaskan Native

Table 5

Correlations for Study 2 Variables of Interest

	1	2	3	4	5	6	7	8	9
Prosocial Responses (1)									
Antisocial Responses (2)	-.22								
Withdrawal Responses (3)	-.18	.61							
Person-Focused (4)	.07	.29	.34						
Behavior-Focused (5)	.17	.09	.09	.28					
Use of Term "Racist" (6)	-.13	.23	.22	.08	.05				
Confronter Race (7)	-.01	.08	.19	.14	.06	.06			
Relationship with Confronter (8)	.09	-.07	-.18	-.07	.04	-.02	-.08		
Participant Gender (9)	-.05	.15	.03	.02	-.04	.03	-.17	-.04	
Participant Age (10)	-.16	-.21	-.16	-.15	-.06	-.18	.03	.10	.01

Note. Lowest $N = 164$. Bolded values are significant at the $p < .05$ level. Use of Term "Racist" coded as 0 = no, 1 = yes. Confronter Race coded as 0 = different from targeted group, 1 = same as targeted group. Relationship with coworker: higher values = closer relationship. Participant Gender coded as 0 = female, 1 = male.

Table 6

Regression Results for Confrontation Responses, Study 2

	Prosocial Responses			Antisocial Responses			Withdrawal/avoidant Responses		
	<i>b</i>	<i>t</i> -value	p-value	<i>b</i>	<i>t</i> -value	p-value	<i>b</i>	<i>t</i> -value	p-value
<u>Control Variable</u>									
Use of Term “Racist”	-.52	-1.92	.057	.59	2.71	.007	.67	2.59	.011
<u>Predictor Variables</u>									
Behavior-Focused	.19	2.18	.031	.01	.09	.928	-.01	-.12	.903
Person-Focused	.03	.34	.731	.20	3.42	<.001	.30	4.21	<.001

Note. *N* = 166

Table 7

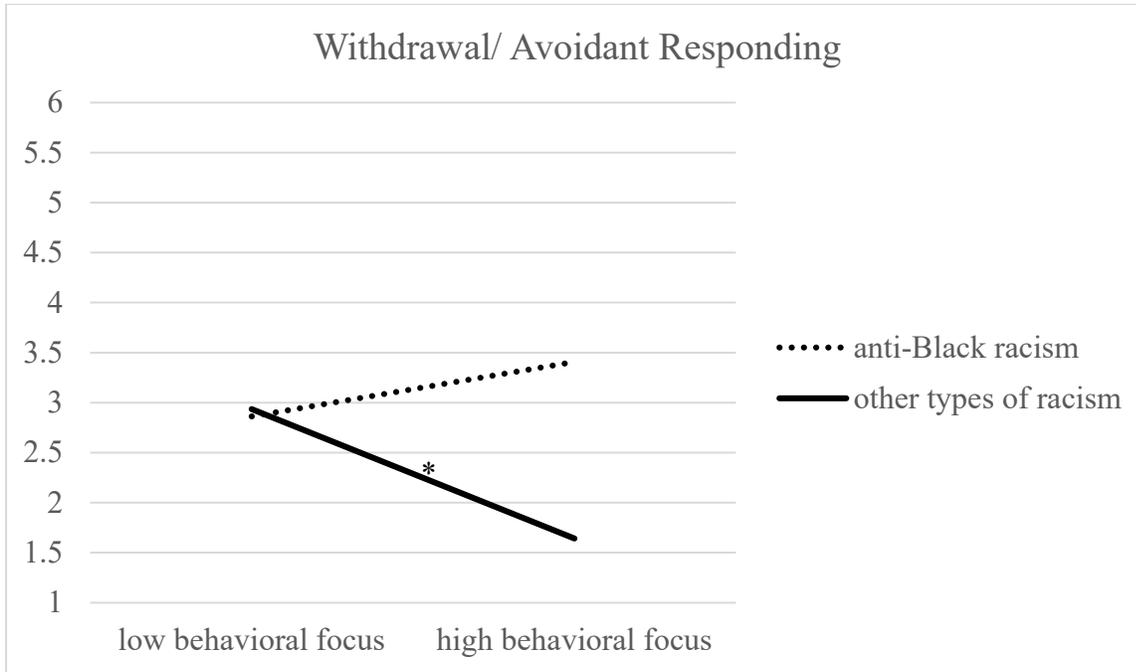
Correlations for Study 3 Variables of Interest

	1	2	3	4	5	6	7	8	9	10	11	12	13
Prosocial Responses (1)													
Antisocial Responses (2)	-.36												
Withdrawal Responses (3)	-.40	.52											
Person-Focused (4)	-.10	.31	.20										
Behavior-Focused (5)	.14	-.16	-.02	.01									
Benevolence Motivation (6)	.54	-.19	-.34	-.10	.18								
Revenge Motivation (7)	-.13	.58	.34	.28	-.24	-.21							
Rel. Repair Beliefs (8)	.14	-.19	-.31	-.18	.03	.09	-.01						
Used the term "racist" (9)	-.01	.26	.12	.15	-.01	.06	.19	-.03					
Social Desirability (10)	.11	-.05	.00	.03	.06	.10	-.10	-.04	-.05				
Confronter Race (11)	-.09	.11	.08	.14	-.09	-.13	.00	-.08	-.14	.04			
Relationship with Confronter (12)	.19	-.06	-.17	-.02	.10	.28	-.03	.24	.03	.05	-.08		
Participant Gender (13)	.08	.00	-.05	.03	-.11	.03	.12	.09	.01	-.01	.02	-.02	
Participant Age (14)	-.11	.01	-.05	.10	-.02	.04	-.01	-.01	-.15	.06	-.08	.11	-.11

Note. Lowest $N = 226$. Bolded values are significant at the $p < .05$ level. Use of Term "Racist" coded as 0 = no, 1 = yes. Confronter Race coded as 0 = different from targeted group, 1 = same as targeted group. Relationship with coworker: higher values = closer relationship. Participant Gender coded as 0 = female, 1 = male.

Figure 1

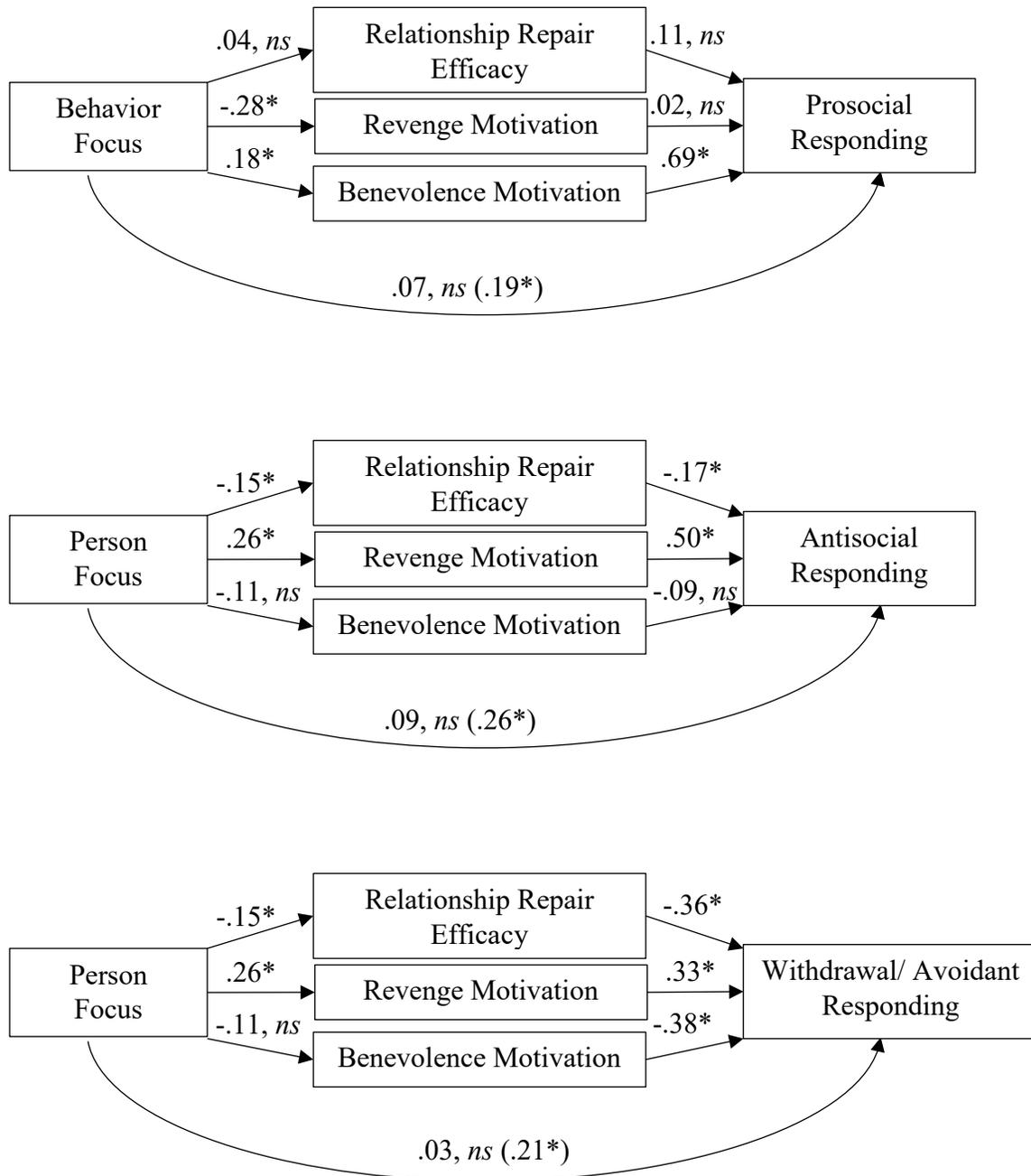
Study 2 interaction between racism type and behavioral focus of confrontation



Note. * indicates significant simple slope at $p < .05$ level.

Figure 2

Hypotheses Testing, Study 3



†Prosocial responding analyses controlled for person focus, using the term “racist” and socially desirable responding; antisocial and withdrawal/avoidant responding analyses controlled for behavior focus, using the term “racist” and socially desirable responding. The values in parentheses are the total effects of X on Y without mediators included.