Forthcoming in *Journal of Experimental Social Psychology*

European Americans' Intentions to Confront Racial Bias: Considering Who, What (Kind), and Why

Riana M. Brown[a], Maureen A. Craig[a], Evan P. Apfelbaum[b]

[a]Department of Psychology, New York University

[b]Questrom School of Business, Boston University

*Please send correspondence to Riana Brown, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003 (e-mail: rmb446@nyu.edu).

Word count = 11,099 (main text, footnotes, and open practices: 10,927; abstract: 172)

Acknowledgements
The authors gratefully acknowledge the support of NSF grant BCS-1823840 awarded to the second author and thank Tessa West for her helpful comments on a draft of this article.

Abstract

Confrontation research has primarily focused on what drives individuals' intentions to confront strangers who express prejudicial attitudes (i.e., *interpersonal bias*; for reviews see Ashburn-Nardo & Karim, 2019; Kawakami et al., 2019). However, bias manifests in multiple forms, including biased policies and institutional practices (i.e., structural bias) or bias perpetrated by close others (e.g., friends), and little is known about what factors impede (or facilitate) intentions to confront these different manifestations of bias. Across three experiments, European Americans reported wanting to confront instances of structural racial bias more than interpersonal racial bias. This was driven by perceptions that the examples of structural bias were more harmful and that confronting would be more effective in changing the perpetrator's behavior, compared with examples of interpersonal bias. Additionally, participants expressed greater intentions to confront friends over strangers (Studies 1-2), due to participants' perceptions that they personally would be effective confronters and that friends would be more receptive. This work provides insight into people's intentions to confront varying manifestations of bias, namely biased structures and close others.

Keywords: structural bias, confrontation, prejudice, close others, racial bias

European Americans' Intentions to Confront Racial Bias: Considering Who, What (Kind), and Why

In 2012, officials in the city of Flint, Michigan introduced a cost-saving policy that switched the water supply from Lake Huron to the Flint River. Given the stark residential segregation in the area, this change in water source exposed citizens from predominantly African American neighborhoods to lead and other contaminants in their tap water. In 2015, pediatrician Dr. Mona Hanna-Attisha, confronted state officials, exposing what she called the environmental and structural racism leading to this crisis (Martin, 2021). Her initial confrontation sparked a national outcry and a series of lawsuits resulting in charges against nine state officials (Brooker, 2021). Given the stark consequences of structural racism (e.g., lead poisoning affecting African American communities), it is crucial to understand when people seek to confront structural problems when they see them. While psychological research has largely examined what drives individuals' decisions to confront strangers who express prejudicial attitudes (i.e., interpersonal bias; for reviews see Ashburn-Nardo & Karim, 2019; Kawakami et al., 2019), as is evident from the situation in Flint, individuals may also seek to confront people enacting biased policies and engaging in biased institutional practices (i.e., structural bias).

In addition to confronting strangers (as occurred in the Flint case), bias may be perpetrated by close others as well (e.g., friends or family; Horowitz et al., 2019). Indeed, the Southern Poverty Law Center (2015) gathered hundreds of stories of everyday bias across the United States and the resultant report highlighted the frequency of close other perpetrators (including friends, neighbors, family, and colleagues). Given that the existing confrontation literature almost exclusively examines when people seek to confront strangers (for a notable exception, see Dickter & Newton, 2013), the present research experimentally manipulates *who*

perpetrates bias and *what kinds* of bias occur to more comprehensively understand confrontation intentions.

Psychological research has established that confrontation (i.e., speaking out against perceived bias) is an effective way to address interpersonally-expressed bias (Chaney & Sanchez, 2018; Chaney et al., 2015; Czopp et al., 2006). Recent work also suggests that European American perpetrators of interpersonal bias are more likely to listen to a European American confronter (vs. an African American confronter; Czopp & Monteith, 2003; Gulker et al., 2013; Rasinski & Czopp, 2010). Thus, the current research broadens the confrontation literature by testing the conditions under which European Americans intend to confront different types of racial bias (interpersonally- or structurally-enacted) perpetrated by different European American actors (friends or strangers).

What Kind: Types of Bias

Different types of bias contribute to racial inequality. Often, psychological research focuses on individual-level racism stemming from negative attitudes and behaviors expressed towards racial minorities (i.e., interpersonal racism; for notable exceptions, see Adams et al., 2008; Nelson et al., 2013; O'Brien et al., 2009; Trawalter et al., 2019). Importantly, racial inequality can also stem from institutional and societal policies, practices, and procedures that detrimentally impact racial minorities (i.e., structural racism; Bonilla-Silva, 1997, 2015). As an example of inequality that may be driven by different modalities of racial bias, a racial disparity exists in healthcare such that African Americans are twice as likely as European Americans to die from heart disease, stroke, and diabetes (National Healthcare Quality and Disparities Reports, 2019). This disparity could stem *both* from interpersonal-level biases in how much time and resources physicians spend on European (vs. African) American patients as well as from

structural factors such as European Americans' greater access to hospitals and pharmacies that are often built in predominantly-European American districts (Williams & Mohammed, 2008). Further, understanding racism through a cultural-psychological lens highlights that while distinguishable, interpersonal and structural forms of racism are inherently intertwined (see Salter et al., 2018). That is, the policies and practices that form structural bias are created and maintained by individuals and these structural aspects, in turn, shape people's experiences and attitudes. Thus, while people may encounter specific instances of bias that falls more neatly into interpersonal or structural racism, these forms of bias work together and impact each other to contribute to inequality.

Previous research has examined people's perceptions and recognition of interpersonal and structural bias (e.g., Nelson et al., 2013; Unzueta & Lowery, 2008). This work typically operationalizes structural racial bias as anonymized racist polices and norms without implicating specific perpetrators. However, given our interest in confrontation intentions (which require a target to potentially confront) and the fact that individuals may be responsible for enacting policies that reproduce racism (Salter et al., 2018), we keep constant in our operationalization of bias a focus on *an individual perpetrator.* Specifically, we examine people's intentions to confront an individual who stereotypes and treats racial minorities negatively (enacting interpersonal racism) or an individual who implements or maintains a biased policy or practice that detrimentally impacts racial minorities (enacting structural racism). From these definitions, structural and interpersonal bias may include both acts of commission (someone personally engages in the biased action) and acts of omission (someone fails to correct someone else's biased action or existing bias; see also Fryberg & Eason, 2017). We examine examples of both, with a particular emphasis on bias involving commission. Overall, the existing confrontation

literature explains what drives peoples' intentions to confront a stranger implementing interpersonal bias (e.g., confronting a biased physician, for review see Ashburn-Nardo & Karim, 2019), but no work (to our knowledge) has examined when and why people seek to *confront an individual* who perpetrates structural bias (e.g., confronting someone who decided where to build new hospitals). As no prior work has compared confrontation of structural (vs. interpersonal) bias, we did not have strong predictions regarding which type of bias may elicit greater intentions to confront.

Who: Perpetrators of Bias

In making confrontation decisions, people may consider not only *what kind* of bias one perpetrates, but also *who* perpetrates it. Work examining internal factors that impede decisions to confront focuses on confrontation of strangers who perpetrate blatant acts of interpersonal bias (e.g., uttering a racial slur; Dickter & Newton, 2013; Kawakami et al., 2009). For example, one may be concerned that a stranger will not change their behavior, if confronted (Rattan & Dweck, 2010). Close others (i.e., friends, family, and colleagues), however, also perpetrate bias (Dickter & Newton, 2013; Guerin, 2005), and the factors that may obstruct (or facilitate) confronting a friend are less well-understood.

The extant literatures examining the "black sheep effect" and ingroup favoritism offer divergent predictions regarding whether one may be more likely to confront a friend or a stranger perpetrating offensive behavior. The "black sheep effect" occurs when individuals monitor and punish ingroup members more than outgroup members for perpetrating norm-violating behavior (Marques et al., 1988; Pinto et al., 2010; Shinada et al., 2004). Research on this effect posits that close ingroup members (i.e., those most central to one's inner circle) are censured more when they betray the group's standards, compared to more peripheral ingroup members or outgroup

members (Pinto et al., 2010). Given that friends and family are particularly central to one's sense of self, and that the behavior of friends and family can reflect on one's own positive image and self-definition (e.g., Andersen & Chen, 2002; Aron & Aron, 1986; Tesser, 1988), this work suggests that European Americans with egalitarian values may seek to confront biased behavior enacted by a European American friend, more than that of a European American stranger ("police close other hypothesis").

Alternatively, ingroup favoritism may involve justifying and favoring the behavior of ingroup members over outgroup members (e.g., Tajfel & Turner, 1979). Recent work in this area reveals that people exhibit favoritism towards central ingroup members (e.g., friends and family) when the hostile intention of an ingroup perpetrator is ambiguous (Campanhã et al., 2011; Otten, 2009; Wang et al., 2016). For example, participants expressed greater tolerance towards an ingroup member if an unfair offer during a monetary allocation game was perceived as unintentional (vs. intentional; Wang et al., 2016). Similarly, individuals reject ambiguously unfair offers less frequently if proposed by a friend, rather than a stranger (Campanhã et al., 2011). Taken together, if a perpetrator's intentions are ambiguous and the ingroup's norms are egalitarian, people may confront a less central ingroup member—a stranger—and justify the behavior of a central ingroup member—a friend. This research leads to the prediction that when witnessing acts that may be considered biased (e.g., subtle racism, Ozier at al., 2019), European Americans may confront a stranger perpetrating bias more than a friend ("rationalize close other hypothesis").

The present research tested these two competing hypotheses: Either European Americans will intend to confront bias perpetrated by a friend more than a stranger ("police close other hypothesis"; Competing Hypothesis 1a) or European Americans will intend to confront bias

perpetrated by a stranger more than a friend ("rationalize close other hypothesis"; Competing Hypothesis 1b).

Why: Determinants of Confronting

What leads people to actively confront bias in whatever form it may emerge? Focusing on the confrontation of interpersonal bias, the Confronting Prejudice Response model (CPR) discusses a series of steps that may occur in the process of deciding whether or not to confront prejudice (see Ashburn-Nardo et al., 2008; Ashburn-Nardo & Karim, 2019). For potential confronters of interpersonal bias, labeling behavior as discrimination and perceiving the behavior as harmful are two important considerations (e.g., Step 1 and 2 of the CPR model, Ashburn-Nardo et al., 2008). To determine if an act qualifies as discrimination (Step 1), people consider perpetrator intent and perceived harm to the target (Swim et al., 2003). Recent work considers the importance of intentionality judgments when determining whether to address racial injustice. Specifically, when considering a case of racial discrimination, intentionality perceptions lead to an increase in desired punishment for the perpetrator (Apfelbaum et al., 2017). Further, European Americans, relative to African Americans, are more influenced by intent and less on harm when judging if acts of ambiguous bias qualify as discrimination (Simon et al., 2019). If labeling behavior as discrimination serves as the first consideration for confrontation, then perceived intent may be particularly pertinent to whether or not European Americans decide to confront.

Research from the perspective of individuals targeted by interpersonal bias (e.g., women confronting sexism) reveals additional considerations for targets' decisions to confront. This work finds that targets are also concerned with the perceived malleability of the perpetrator (e.g., will perpetrators change and listen?; Kaiser & Miller, 2004; Rattan & Dweck, 2010; Rattan, 2019), feeling personally efficacious as a confronter (e.g., perceiving perpetrators will listen to

them personally; Good et al., 2012; Rattan, 2019), and perceived physical costs (e.g., safety concerns; Ayres et al., 2009) and affective costs (e.g., being labeled as a complainer, Eliezer & Major, 2012; Good et al., 2012; Kaiser & Miller, 2004). Although primarily studied in the target confrontation literature, these considerations may also be present for confronters not belonging to the targeted group.

Importantly, while the confrontation literature discusses determinates to confront interpersonal bias, the research on collective action and allyship examines when people engage in actions (e.g., attending protests, volunteering for activist organizations) to address structural inequities. Aligned with the target confrontation literature, the perception that engaging in collective action will instill change (Hornsey et al., 2006) and beliefs that the group is able to effectively achieve their goals (i.e., group-based efficacy; van Zomeren, 2013; van Zomeren et al., 2008) motivates engagement in collective action to address social inequality. Further, advantaged group members support collective action to address structural bias and seek to act when they perceive the status quo as unjust (e.g., Dietze & Craig, 2020) and see inequalities as impactful and unfair (e.g., perceived harm, Iyer & Ryan, 2009; for reviews see Craig et al., 2020; Louis et al., 2019; Radke et al., 2020). Taken together, this research suggests that perceiving that confrontation will effectively enact change (i.e., that the perpetrator will change and listen if confronted), perceiving oneself as an effective confronter, and perceiving harm to targets of bias will lead European Americans to seek to confront bias.

The present research examined whether the various mechanisms highlighted by the confrontation of interpersonal bias and collective action literatures vary depending on what kind of bias a perpetrator commits and one's relationship to the perpetrator. We predict (Hypothesis 2) that European Americans will support confronting someone for perpetrating bias when they

perceive greater intent, greater perpetrator malleability, a greater sense of personal efficacy, less physical costs, less affective costs, and greater harm.

Overview of Experiments

Three experiments examined European Americans' perceptions of and intentions to confront different forms of racial bias directed at African Americans (interpersonal or structural bias; Studies 1-3) perpetrated by different actors (strangers or friends; Hyp. 1 tested in Studies 1-2). Across studies, we also examined the underlying processes that drive intentions to confront (e.g., perceptions of intentionality, harm, and perpetrator malleability; Hyp. 2) different forms of bias enacted by different actors. Overall, these studies empirically tested what leads European Americans to seek to confront different types of common manifestations of racial bias (i.e., interpersonal or structural) that are perpetrated by people who are more central (i.e., friends) or peripheral (i.e., strangers) in their social circle. In these studies, we report all measures, manipulations, and exclusions (see footnotes for additional measures not examined in the main text).

Study 1

Study 1 focuses on the role intentionality plays in intentions to confront different forms of bias perpetrated by different actors. Given that European Americans, relative to African Americans, are more influenced by intent when judging if acts of ambiguous bias qualify as discrimination (Simon et al., 2019), we measure how perceived intentionality may affect confrontation intentions. European Americans read vignettes involving structural bias or interpersonal bias ostensibly perpetrated by either their friend or a stranger and rated the intentionality of the behavior and their confrontation intentions for each scenario. Given that we were interested in confrontation intentions, which require a target to confront, each vignette

involved an individual perpetrator. As a result, the interpersonal bias vignettes depicted an individual either treating African Americans negatively or allowing African Americans to be treated negatively, while the structural bias vignettes depicted an individual implementing or maintaining a biased policy or practice that negatively impacted African Americans. The vignettes also described relatively ambiguous instances of racial bias as they did not explicitly indicate the exhibited behavior to be racially motivated. Lastly, the vignettes included examples of both commission (perpetrator personally implements the biased action) and acts of omission (perpetrator fails to correct someone else's or existing biased action).

Given our competing hypotheses, we expected that participants may either report greater intentions to confront a friend perpetrating bias than a stranger (Hyp. 1a, "police close other hypothesis") or more intentions to confront a stranger than a friend (Hyp. 1b, "rationalize close other hypothesis"). Based on prior work (Blodorn et al., 2012), we also predicted that participants would rate structurally-enacted bias as less intentionally perpetrated compared to acts of interpersonal bias and explored whether this would influence which type of bias (structural or interpersonal) participants sought to confront. Following this prior work, if interpersonal bias is perceived as more intentionally perpetrated than structural bias, European Americans may report greater intentions to confront interpersonal bias compared to structural bias.

Method

Participants. Two hundred and sixteen European Americans (41% female, age range: 18-74, M_{age}=36.25, SD_{age}=11.93) were recruited from MTurk.com in exchange for $1. Sample size was determined using G*Power to detect a small-medium effect size (η_p^2=.035) for a 2x2

between-subjects interaction effect yielding a required sample size of 219 to achieve 80% power. Sample size was determined before any data analysis.

Materials and measures

Perpetrator closeness manipulation. All participants were randomly assigned to read three vignettes depicting ambiguous instances of bias. Two of the vignettes involved acts of commission (perpetrator engages in the biased action personally) and one involved an act of omission (failing to correct someone else's or existing biased action; see Appendix Table 1 for vignettes)[1]. To manipulate the participants' closeness to the perpetrator, participants were asked to imagine that either their closest friend perpetrated the described behaviors or to imagine they did not know the person in the vignettes (adapted from Heiphetz & Craig, 2020). In the friend condition, participants were told that "some of these behaviors may be very unlike your friend, but please do your best to imagine these hypothetical vignettes…". They were then prompted to write down the first name of their friend (any gender) and indicate their friend's gender pronoun preference (*he, she,* or *they*). The name of their friend and the given pronoun were piped into all vignettes. In the stranger condition, participants read that the behaviors were perpetrated by a stranger named Jamie and were asked to provide their own gender pronoun preference (*he, she,* or *they*). The stranger's gender matched participants' gender.

Bias type manipulation. To manipulate the type of bias, half of participants were randomly-assigned to view vignettes involving interpersonally-enacted bias (adapted from Corning & Bucchianeri, 2010). For example, "Yesterday, [friend's name/Jamie] was on the downtown bus. After a few stops, a Black couple boarded and sat down next to Jamie. Shortly after the couple sat down, [he/she/they] got up, walked down the aisle, and held a handrail. [Friend's name/Jamie] did not get off at the next stop". The other half of participants viewed

structural-based vignettes, which depicted someone acting in a way that would disproportionately impact African Americans through creating a policy (e.g., voting for a new town law) or maintaining a common societal practice (e.g., noticing racial pay gaps). For example, "[Friend's name/Jamie] went to the local town hall. During the session, the council proposed a motion that would divert money from local transportation initiatives to predominantly-Black communities exposed to environmental hazards (e.g., lead paint and diesel fumes). During the meeting, [friend's name/Jamie] suggested that there was nothing we could do about this problem and that the money would be more useful for education". We created the structural vignettes using Unzueta and Lowery's (2008) scenarios as guides.

Confrontation intentions. Participants indicated how much they thought the perpetrator (stranger or friend) should be confronted for their behavior (1=*not at all*, 6=*very much so*). Confrontation intentions for the three vignettes were averaged (3-items; α=.74) with higher scores indicating greater support for confrontation.

Perceived intentionality. Participants indicated the extent to which they thought the behavior of the stranger or their friend (depending on condition) was intentional (1=*not at all,* 6=*very much so*). Perceived intent for the three vignettes were averaged (3-items; α=.72) with higher scores indicating greater perceived intent.

Manipulation check. Participants were asked about the nature of racial bias to assess if exposure to the different forms of bias influenced perceptions that racism stems from interpersonal or structural factors (adapted from Craig et al., 2020; O'Brien et al., 2009). Two questions assessed perceptions that racism is interpersonally-driven (e.g., "Racism is primarily caused by racist individuals who have negative attitudes toward racial minorities"; averaged 2-item interpersonal racism index; r=.54, p<.001) and two assessed perceptions that racism is

structurally-driven (e.g., "Racism is primarily caused by institutional practices and structural factors (e.g., laws, policies) that disadvantage racial minorities," 1=*strongly disagree*, 7=*strongly agree*; averaged 2-item structural racism index; r=.78, p<.001). These indices were uncorrelated with one another (r=.05, p=.428). Lastly, a difference score (structural index-interpersonal index) was created with positive scores indicating perceptions that racism is driven by structural causes more than interpersonal causes (and *vice versa* for negative scores)[2].

Procedure. After providing informed consent, participants read three counterbalanced ambiguously-biased vignettes (either describing interpersonal or structural bias) and imagined the behavior was either perpetrated by a friend or a stranger. Each scenario was followed by questions measuring perceived intentionality and confrontation intentions. Participants also indicated the race and gender of the imagined perpetrator, answered the manipulation check questions[3], and demographic questions. Finally, participants were thanked and debriefed.

Results

To keep constant that all participants considered a European American perpetrator, we excluded 17 participants who indicated that the imagined actor in the vignettes was not European American, yielding a final sample of 199 (47 Interpersonal-stranger condition, 58 Interpersonal-friend condition, 58 Structural-stranger condition, 36 Structural-friend condition)[4]. The final sample has 80% power to detect small-medium effects (Cohen's f =.20). See Table 1 for correlations and Table 2 and Table 3 for descriptive statistics of the manipulation check measures and main dependent measures, respectively. We conducted a series of analyses of variance (ANOVAs), followed by planned contrasts to assess the differences among the experimental conditions.

Manipulation check. We first tested if participants' perceptions that racism was driven by interpersonal or structural factors differed by the type of bias highlighted in the vignettes, perpetrator closeness, and their interaction. The main effect of perpetrator closeness and interaction between bias type and perpetrator closeness were nonsignificant (*ps*>.206). Importantly, suggesting that the bias type manipulation was effective, a main effect of bias type emerged, $F(1, 195)=5.16$, $p=.024$, $\eta_p^2=.03$. Participants exposed to the structural vignettes ($M=0.20$, $SD=1.83$) perceived racism to be driven more by biased structures (vs. negative attitudes) compared with participants exposed to the interpersonal vignettes ($M=-0.44$, $SD=2.11$), regardless of whether a friend or stranger perpetrated (see Table 2).

Confrontation intentions. We next examined ratings of confrontation intentions. Consistent with the "police close other hypothesis", a main effect of perpetrator closeness emerged, $F(1, 195)=6.36$, $p=.012$, $\eta_p^2=.03$. Participants considering their friend ($M=3.97$, $SD=1.32$) perpetrating bias reported greater support for confrontation than participants considering a stranger ($M=3.55$, $SD=1.42$), regardless of bias type. Furthermore, a main effect of bias type emerged, $F(1, 195)=7.61$, $p=.006$, $\eta_p^2=.04$; people indicated that the structural vignettes ($M=4.12$, $SD=1.06$) should be confronted more than participants reading interpersonal vignettes ($M=3.52$, $SD=1.53$), regardless of who perpetrated (see Figure 1 and Table 3). No interaction emerged between relationship type and bias type, $F(1, 195)=2.22$, $p=.138$, $\eta_p^2=.01$.

Perceived intentionality. We next examined ratings of perceived intentionality. Consistent with prior research (Blodorn et al., 2012), a significant main effect of bias type emerged. Regardless of perpetrator closeness, participants exposed to structural bias vignettes rated the actor's behavior as less intentional ($M=4.38$, $SD=1.02$) than those considering interpersonal bias vignettes ($M=4.75$, $SD=1.14$), $F(1, 195)=6.23$, $p=.013$, $\eta_p^2=.03$. There were

no significant differences in perceived intentionality by perpetrator closeness nor by the bias type X perpetrator closeness interaction (*ps*>.750). Overall, acts of interpersonal bias were perceived as more intentionally perpetrated compared to acts of structural bias.

Mediation. Given that perceived intentionality is an important precursor for recognizing occurrences as reflecting bias (Simon et al., 2019; Swim et al., 2003), intentionality judgments may similarly influence support for confrontation. To investigate whether differences in perceived intentionality may underlie the effects of bias type on confrontation intentions, we conducted a mediation analysis (with 10,000 bootstrap samples; Hayes, 2017, model 4). The analysis revealed that perceived intentionality statistically mediated the effect of bias type on confrontation intentions (Indirect effect: -0.12, 95% CI[-0.28, -0.03]; see Figure 2). European Americans perceived interpersonally-biased actions as more intentional and that more intentional behaviors should be confronted more. The direct and total effects of bias type, however, indicated greater support for confronting structural bias (over interpersonal bias), suggesting that people may weigh other factors more than perceived intentionality in their confrontation decision-making process.

Discussion

Supporting the "police close other hypothesis," people rated that ingroup friends perpetrating bias should be confronted more than ingroup strangers. This is consistent with the idea that European Americans may confront close others as a way of restoring positive self- and group-image (Pinto et al., 2010). Further, European Americans reported that structural bias should be confronted *more* than interpersonal bias. This finding is notable given that interpersonal bias was perceived as more intentional than structural bias (consistent with prior work; Simon et al., 2012) and perceived intent predicted greater support for confrontation. This

suggests that other, unmeasured factors underlie participants' general support for confronting structural bias more than interpersonal bias (and particularly a stranger perpetrating structural bias as suggested by the simple effects, see Figure 1 and Table 3).

Study 2

Study 2 sought to replicate Study 1 and measure additional factors that may underlie the confrontation of different perpetrators and different forms of bias. Guided by the target confrontation and collective action research, people may weigh if they personally are the right person to confront (i.e., personal efficacy, Good et al., 2012; van Zomeren, 2013) and if confrontation will effectively change the perpetrator's behavior (i.e., perceived perpetrator malleability, Hornsey et al., 2018; Rattan, 2019; Rattan & Dweck, 2010). A potential confronter may also be concerned with their physical safety during the confrontation and their potential feelings of distress (i.e., potential physical and affective costs; Kaiser & Miller, 2004; Shelton & Stewart, 2004). These factors may influence intentions to confront more generally or may vary depending on *who* is perpetrating and the *type* of bias manifested. Study 2 tested if these concerns explain why European Americans seek to confront friends more than strangers and confront structural bias more than interpersonal bias.

Study 2 also incorporated a mixed-design to increase power to detect an interaction (Charness et al., 2012). Participants were randomly assigned to imagine either a friend or a stranger (between-subjects factor) perpetrating subtly-biased behaviors (both interpersonal and structural, as a within-subjects factor). To better capture participant's confrontation intentions, participants in Study 2 indicated their desire to personally confront in addition to their perception that the perpetrator should be confronted more generally. Based on the results of Study 1 and the "police close other hypothesis" (Hyp. 1a), we predicted that participants would support

confronting a friend more than a stranger for perpetrating interpersonal bias. Also based on the results of Study 1, we expected that participants would support confronting someone perpetrating structural bias over interpersonal bias. We also explored whether potential mechanisms identified from prior work (perceived intentionality, personal efficacy, perpetrator malleability, physical costs, and affective costs) may drive support of confronting different types of bias (structural vs. interpersonal) perpetrated by different actors (friends vs. strangers).

Method

Participants. Three hundred and thirty-two European Americans (51% female, age range: 18-78, M_{age}=34.18, SD_{age}=11.43) were recruited from Prolific Academic in exchange for $1.65. We sought to collect a sample of at least 278 participants, to achieve 80% power to detect a small effect (f=.10) in a 2x2 mixed-design interaction. Sample size was determined before any data analysis.

Materials and measures

Perpetrator closeness manipulation. As in Study 1, participants were randomly assigned to either imagine that their closest friend or a stranger committed a series of biased behaviors.

Bias type manipulation. Participants read the same bias vignettes used in Study 1. Importantly, in Study 2, bias type was a within-subjects factor, and participants read all six vignettes (three depicted interpersonal bias and three depicted structural bias). We counter-balanced the order of presented vignettes and block-presented the three interpersonal bias vignettes together and the three structural bias vignettes together.

Manipulation check. Because bias type was a within-subjects manipulation, we included a manipulation check after each scenario by asking the extent to which the vignette behavior depicted an example of interpersonal bias or structural bias. To assess interpersonal bias

perceptions, participants were asked how much the perpetrator's behavior was an example of someone who has negative attitudes toward racial minorities (1=*not at all*, 6=*very much so*). We created indices for each bias type by averaging the ratings of the three interpersonal vignettes (3-items; α=.66) and the three structural vignettes (3-items; α=.71) separately. Higher scores indicate greater perceptions that the vignettes depicted interpersonal bias.

To assess structural bias perceptions, participants indicated how much the perpetrator's behavior depicted common practices and norms that disadvantage racial minorities (1=*not at all*, 6=*very much so*). We created indices for each bias type by averaging the ratings of the three interpersonal vignettes (3-items; α=.78) and the three structural vignettes (3-items; α=.83) separately. Higher scores denote greater perceptions that the vignettes depicted structural bias. We again created differences scores with positive values indicating perceptions that the vignettes depicted structural bias more than interpersonal bias and negative scores indicating perceptions of more interpersonal than structural bias.

Confrontation intentions. To measure confrontation intentions, participants indicated if the perpetrator's behavior should be confronted, as in Study 1, and how much they would confront the perpetrator themselves (2-items; 1=*not at all*, 6=*very much so*). We created two confrontation intention indices by averaging the three interpersonal vignettes (6-items; α=.86) and the three structural vignettes (6-items; α=.87) separately. Higher scores denote greater confrontation intentions.

Perceived intentionality. Perceived intentionality was measured with two items: the same item from Study 1 measuring perceptions that the perpetrator's behavior was intentional and a new item measuring perceptions that the perpetrator's behavior was accidental (reverse-coded; 1=*not at all*, 6=*very much so*). We created two perceived intentionality indices by averaging the

three interpersonal vignettes (6-items; α=.72) and the three structural vignettes (6-items; α=.73) separately. Higher scores indicate greater perceived intentionality.

Personal efficacy. To measure feelings of personal efficacy in confronting, participants indicated their perceptions that they personally could effectively address the situation and that they are the right person to confront the perpetrator (2-items; 1=*not at all,* 6=*very much so*). We created two personal efficacy indices by averaging the three interpersonal vignettes (6-items; α=.93) and the three structural vignettes (6-items; α=.93). Higher scores depict greater perceived personal efficacy.

Perpetrator malleability. To assess whether people perceived the perpetrator would change their behavior after being confronted, participants rated how much confronting would change the perpetrator's behavior and if the perpetrator would listen if confronted (2-items; 1=*not at all,* 6=*very much so*). We created two perpetrator malleability indices by averaging the three interpersonal vignettes (6-items; α=.89) and the three structural vignettes (6-items; α=.90). Higher scores denote greater perceived perpetrator malleability.

Perceived costs. To measure perceived physical costs, participants indicated if they were afraid of how the perpetrator would act if confronted and how much they thought the perpetrator might act aggressively if confronted (2-items). To measure affective costs, participants rated how awkward and how uncomfortable it would be to confront the perpetrator (2-items; 1=*not at all,* 6=*very much so*). Items for the three interpersonal vignettes (physical costs: 6-items; α=.92; affective costs: 6-items; α=.95) and the three structural vignettes (physical costs: 6-items; α=.92; affective costs: 6-items; α=.96) were averaged to create indices, with higher scores representing greater perceived physical or affective concerns.

Procedure. After providing informed consent, participants read six vignettes (describing both interpersonal and structural bias) perpetrated by either a friend or a stranger. Each scenario was followed by manipulation check questions and questions assessing perceived intentionality, confrontation intentions, personal efficacy, perpetrator malleability, and physical and affective costs. Participants also indicated the imagined perpetrator's race and gender and answered demographic questions. Finally, participants were thanked and debriefed.

Results

We excluded 25 participants who indicated that the imagined actor was non-European American yielding a sample of 307 participants (155 stranger condition, 152 friend condition). The final sample has 80% power to detect small effects (f =.09) in a within-between interaction. See Table 1 for correlations and Table 2 and Table 3 for descriptive statistics of the manipulation check measures and main dependent measures, respectively (see supplemental Table for additional correlations within bias conditions). We conducted a series of mixed-design ANOVAs (bias type as the within-subjects factor and relationship closeness as the between-subjects factor) followed by planned contrasts testing the differences among the experimental conditions (Bonferroni corrections were used for all pairwise comparisons).

Manipulation check. We first tested if participants perceived the vignettes in ways consistent with the definitions of interpersonal bias and structural bias. Assessing the difference score (structural index-interpersonal index), a main effect of bias type emerged, $F(1, 305)=135.50$, $p<.001$, $\eta_p^2=.31$. Aligned with predictions, participants perceived the structural vignettes (M=0.53, SD=0.95) were driven more by biased structures (vs. individuals with negative attitudes) compared with the interpersonal vignettes (M=-0.26, SD=1.03), regardless of whether it was a friend or stranger perpetrating (see Table 2). Unexpectedly, a main effect of

perpetrator closeness also emerged, $F(1, 305)=4.94$, $p=.027$, $\eta_p^2=.02$; participants imagining a friend ($M=0.23$, $SD=0.74$) rated the actor as a better example of someone acting based on biased structures (vs. having negative attitudes) than those imagining a stranger ($M=0.04$, $SD=0.74$), regardless of bias type. In other words, the stranger was seen as a better example of someone who acted with negative attitudes (vs. acting based on biased structures) compared to the friend. There was no interaction between bias type and perpetrator closeness, $F(1, 305)=1.24$, $p=.265$, $\eta_p^2=.00$.

Confrontation intentions. We next examined confrontation intentions. Consistent with Study 1, main effects of perpetrator closeness, $F(1, 305)=7.90$, $p=.005$, $\eta_p^2=.02$, and bias type emerged, $F(1, 305)=55.28$, $p<.001$, $\eta_p^2=.15$. Participants indicated that examples of structural bias should be confronted more than interpersonal bias, regardless of whether one imagined a friend or stranger perpetrator. Aligned with the "police close others hypothesis", participants imagining a friend rated that the behavior should be confronted more than participants imagining a stranger, regardless of bias type. Furthermore, likely due to the increased statistical power in Study 2, these main effects were qualified by a significant interaction, $F(1, 305)=4.69$, $p=.031$, $\eta_p^2=.02$. Consistent with patterns revealed in Study 1, participants reported the lowest intentions to confront a stranger committing interpersonal bias, compared with a stranger committing structural bias or a friend perpetrating either type of bias ($ps<.001$, see Table 3).

Perceived intentionality. We next tested if perceptions of intentionality differed depending on who perpetrated the behavior and the type of biased action committed. Consistent with Study 1, a main effect of bias type on intentionality ratings emerged, $F(1, 305)=28.59$, $p<.001$, $\eta_p^2=.09$; participants rated structural bias ($M=4.61$, $SD=0.96$) as less intentional than interpersonal bias ($M=4.90$, $SD=0.92$), regardless of the imagined perpetrator. Unlike Study 1, a

main effect of perpetrator closeness emerged, $F(1, 305)=10.85$, $p=.001$, $\eta_p^2=.03$; participants imagining a stranger ($M=4.90$, $SD=0.68$) rated the actor as more intentional, compared to participants imagining their friend ($M=4.60$, $SD=0.89$), regardless of bias type. There was no reliable interaction between bias type and perpetrator closeness, $F(1, 305)=0.46$, $p=.500$, $\eta_p^2=.00$.

Personal efficacy. When assessing ratings of personal efficacy in confronting, main effects of perpetrator closeness, $F(1, 305)=69.10$, $p<.001$, $\eta_p^2=.18$, and bias type emerged, $F(1, 305)=5.46$, $p=.020$, $\eta_p^2=.02$, qualified by a significant interaction, $F(1, 305)=5.70$, $p=.018$, $\eta_p^2=.02$. Participants imagining their friend reported feeling more personally effective at confronting compared to participants imagining a stranger, regardless of bias type. Further, simple effects reveal that for participants imagining a stranger, they reported feeling more efficacious if confronting structural bias compared to interpersonal bias ($p<.002$, see Table 3). In contrast, for participants imagining a friend, there were no differences in reported feelings of efficacy depending on bias type ($p>.970$, see Figure 4).

Perpetrator malleability. When assessing perceptions that the perpetrator would change if confronted, we found main effects of perpetrator closeness, $F(1, 305)=70.66$, $p<.001$, $\eta_p^2=.19$, and bias type, $F(1, 305)=65.87$, $p<.001$, $\eta_p^2=.18$. Consistent with the confronting intentions measure, participants considering their friend reported they would listen more and change if confronted compared to participants imagining a stranger, regardless of the type of bias perpetrated. Participants also rated that the someone perpetrating structural bias would listen more and change if confronted compared to interpersonal bias, regardless of whether the perpetrator was a friend or stranger. Additionally, these effects were qualified by a marginally-significant interaction, $F(1, 305)=3.73$, $p=.054$, $\eta_p^2=.01$, such that participants reported the friend

perpetrating structural bias would listen and change the most, while the stranger perpetrating interpersonal bias would listen and change the least (*ps*<.001, see Table 3 and Figure 4).

Perceived costs. When evaluating perceived physical costs to confronting, a main effect of perpetrator closeness, $F(1, 305)=50.83$, $p<.001$, $\eta_p^2=.14$, and a marginal main effect of bias type emerged, $F(1, 305)=3.10$, $p=.079$, $\eta_p^2=.01$. There was no interaction effect, $F(1, 305)=1.02$, $p=.312$, $\eta_p^2=.00$. Participants imagining a stranger ($M=2.66$, $SD=1.19$) reported that the perpetrator would act more aggressively if confronted than participants imagining a friend ($M=1.78$, $SD=0.96$), regardless of the type of bias. Participants also indicated that someone enacting structural bias ($M=2.18$, $SD=1.21$) would act somewhat less aggressively compared to someone enacting interpersonal bias ($M=2.26$, $SD=1.27$), regardless of perpetrator closeness.

When assessing reported affective costs to confronting, main effects of perpetrator closeness, $F(1, 305)=35.43$, $p<.001$, $\eta_p^2=.10$, and bias type emerged, $F(1, 305)=5.24$, $p=.023$, $\eta_p^2=.02$. Participants imagining a stranger ($M=4.06$, $SD=1.38$) rated it would be more awkward and uncomfortable to confront the perpetrator, compared with participants imagining their friend ($M=3.06$, $SD=1.54$), regardless of bias type perpetrated. Additionally, participants reported that it is less awkward to confront someone enacting structural bias ($M=3.50$, $SD=1.60$) compared with someone enacting interpersonal bias ($M=3.63$, $SD=1.65$), regardless of the imagined perpetrator. The interaction did not reach statistical significance, $F(1, 305)=2.53$, $p=.113$.

Overall, when assessing perceived costs of confronting, participants who imagined a friend reported fewer concerns about aggressive behavior and awkwardness. Similarly, participants reported fewer concerns when exposed to structural bias, compared to interpersonal bias.

Mediation effect. To assess whether these factors (i.e., perceived intentionality, personal efficacy, perpetrator malleability, physical costs, and affective costs) statistically account for the significant interaction between bias type (a within-subjects factor) and perpetrator closeness (between-subjects factor) on confrontation intentions, we conducted a parallel mediation analysis (with 10,000 bootstrap samples, Hayes, 2017, model 4). Relationship type was used as the independent variable and we created difference scores of the within-subjects manipulation (structural-interpersonal) for each of the mediators and the dependent variable. Perceived personal efficacy and perpetrator malleability statistically mediated the interactive effect of perpetrator closeness (friend vs. stranger) by bias type (structural vs. interpersonal) on intentions to confront (personal efficacy indirect effect: 0.16, 95% CI [0.03, 0.32]; perpetrator malleability: 0.06, 95%CI [0.01, 0.14]; see Figure 5). This reveals that the degree to which participants support confronting a friend's (vs. stranger's) biased actions, dependent on the type of bias (structural or interpersonal) perpetrated, was driven by participants' perceptions that they would personally be effective confronters and that the perpetrator would be more likely to change. Perceived intentionality (0.02, 95% CI [-0.04, 0.09]), physical costs (-0.01, 95% CI [-0.06, 0.01]), and affective concerns (-0.01, 95% CI [-0.04, 0.02]) did not reliably mediate[5].

Discussion

Consistent with the "police close others hypothesis" and results of Study 1, European Americans expressed more support for confronting a friend than a stranger. This appears to be due to perceptions that a friend would change if confronted and participants themselves would be effective confronters. New to Study 2 and consistent with the hypothesis that people rationalize the behavior of close ingroup members, participants considering a stranger perpetrator reported that the individual acted more intentionally and their behavior reflected someone with negative

attitudes towards racial minorities, compared with participants considering perpetrating friends. This may suggest that people both "police close others" and "justify close others," but under different circumstances. For example, people may justify their friends' behavior by downplaying intent, while concurrently seeking to police their friends' behavior through confrontation. This interpretation is consistent with recent work considering how protecting behavior is distinct from disciplining behavior. Specifically, people anticipate protecting close others who commit a moral transgression (e.g., by lying to the police), while also seeking to discipline them in private (e.g., talking alone with the transgressor; Weidman et al., 2019). Future work can further test the intriguing possibility that people may prefer different forms of confrontation for close others (private confrontation), compared with strangers (public confrontation).

Results of Study 2 also indicate that the effect of bias type on confronting intentions is dependent on the relationship closeness with the perpetrator. Likely due to the increased power of Study 2's mixed-design, an interaction between perpetrator closeness and bias type emerged—participants expressed the most interest in confronting a friend perpetrating structural bias and the least interest in confronting a stranger perpetrating interpersonal bias (see Figure 2). Replicating Study 1, European Americans indicated the lowest intentions to confront a stranger committing interpersonal bias, compared with a stranger committing structural bias or a friend perpetrating either type of bias ($p<.005$; see Figures 1 and 3). These results suggest that the perceived impediments to confronting strangers (particularly perceptions that one is unable to effectively confront and that the perpetrator will not change) appear to diminish if considering strangers perpetrating structural bias, which facilitates enhanced confronting intentions. This is notable, given that the vast majority of bias confrontation theories and literature consider this

form of bias (interpersonal bias committed by a stranger; e.g., Ashburn-Nardo & Karim, 2019) and highlights the importance of considering additional kinds of perpetrators and forms of bias.

Aligned with past research on confrontations of interpersonal bias (Rattan, 2019; Rattan & Dweck 2010) and motivations to engage in collective action (Hornsey et al., 2018; van Zomeren, 2013), perceptions of oneself as an effective confronter and that the perpetrator would change and listen if confronted drove participants' support for confronting. Perhaps people feel more effective and that the perpetrator will listen during a confrontation of structural bias if they perceive that structural racism involves 'the system' and not the perpetrator's individual bad moral character. Future research can assess who people feel are to blame for different manifestations of bias and how that influences confrontation intentions.

However, perceptions of affective costs and physical safety concerns did not reliably explain support for confrontation of different perpetrators enacting different instantiations of bias. While perceived affective and physical costs deter target confrontations of more overt bias (e.g., Good et al., 2012; Ayres et al., 2009), the present work suggests these concerns may be less central for European American observers of more ambiguously-motivated racial bias. We sought to further understand what underlying motivations influence confrontation intentions in Study 3.

Study 3

Study 3 was a preregistered experiment aimed to address several limitations of Studies 1-2. First, Study 3 assessed an important factor that may differ between perceptions of interpersonal and structural racial bias—perceived harm. Drawing on previous theory and research, the CPR model (Ashburn-Nardo et al., 2008) posits that perceived harm to the target of bias is one of the prerequisites to confronting interpersonal bias (Step 2). Further, the collective action work finds that perceiving harm is necessary for advantaged group members to engage in

collective action to address structural bias (see Craig et al., 2020 for review). Considering that structural bias by definition has a wide-reaching negative impact (Bonilla-Silva, 1997, 2015), people may perceive structural bias as more harmful; this may explain the overall tendency to report greater support for confronting structural, than interpersonal, bias.

Additionally, we revised the experimental materials for Study 3 to meticulously match the setting of the bias, whether the perpetrator engaged in commission or omission, and whether the person(s) targeted noticed they were the target of biased behavior across the conditions (see Appendix Table 2 for the vignettes). Participants read 6 vignettes depicting either interpersonal or structural bias and reported their perceptions that the behavior was harmful, that the perpetrator acted with intent, feelings of efficacy, if the perpetrator would change, the affective and physical costs, and confrontation intentions. To be consistent with the current focus of the confronting literature and to isolate the effect of bias type, we sought to focus on the effect of bias type (structural or interpersonal) for a stranger perpetrating bias.

Based on the results of Studies 1-2, we predicted that perceived intentionality would correlate with confrontation intentions but not account for the overall expected tendency of participants to support confronting structural bias over interpersonal bias. Rather, based on the collective action literature, we expected European Americans to support confronting someone enacting structural bias over interpersonal bias due to perceptions that structural bias was more harmful, that they would be effective confronters, and that the perpetrator would change.

Method

Participants. Five hundred and fifty-two European Americans (54% female, age range: 18-75, M_{age}=38.46, SD_{age}=12.13) were recruited from MTurk.com in exchange for $1. As described in the preregistration document, we sought to collect a sample of at least 501

participants for this study (for more details on sample size and power, see http://aspredicted.org/blind.php?x=fm6gx6).

Materials and measures

Bias type manipulation. Participants read six vignettes, the content of which was randomly assigned. Half of participants were randomly assigned to read vignettes describing interpersonal bias, the other half read vignettes depicting structural bias. We adjusted the experimental materials to closely match the setting of the bias (e.g., if the perpetrator witnessed or instigated bias, if the person(s) targeted noticed the behavior) across conditions (see Appendix Table 2 for all vignettes).

Manipulation check. All participants were asked the same questions from Study 2 assessing participants' perceptions that the vignettes depicted examples of interpersonal bias (6-items; α=.78) and structural bias (6-items; α=.85). Again, we computed the difference score (structural index-interpersonal index) such that positive scores indicate more structural perceptions (than interpersonal perceptions) and *vice versa* for negative scores[6].

Confrontation intentions. The same two items from Study 2 were used to assess support for confrontation and averaged to create a confrontation intentions index (12-items; α=.86).

Perceived intentionality. The same items from Study 2 were used to assess perceived intentionality of the perpetrator and averaged to create a perceived intentionality index (12-items; α=.80).

Perceived harm. To assess perceived harm, we included two items representing general harm and two items on numerical impact (Bonilla-Silva, 2015). Participants reported the degree to which the perpetrator's behavior was harmful and had a negative impact (1= *not at all*, 6=*very much so*) as well as perceptions that people would be affected by the perpetrator's behavior (1=

not at all, 6=*very much so*) and how many people would be affected by the behavior (1=*0 people,* 6= *50 or more people*). These items were standardized (due to the different scale labels) and averaged to create a perceived harm index (24-items; α=.93) with higher scores indicating greater perceived harm[7].

Personal efficacy. The same two items from Study 2 were used to measure perceptions that participants themselves would be effective confronters. Items for each of the six vignettes were averaged to create a personal efficacy index (12-items; α=.93).

Perpetrator malleability. The same two items from Study 2 were used to measure perceptions that the perpetrator would change if confronted. Items for each of the six vignettes were averaged to create a perpetrator malleability index (12-items; α=.87).

Perceived costs. The same items from Study 2 were used to measure physical costs and affective costs of confronting. Items for each of the six vignettes were averaged to create a physical cost index (12-items; α=.93) and an affective cost index (12-items; α=.95).

Procedure. After providing informed consent and answering initial demographic questions (e.g., race, gender), participants read six counter-balanced vignettes (describing either interpersonal or structural bias) perpetrated by a stranger. Each scenario was followed by manipulation check questions, mediator measures (perceived intentionality, perceived harm), the main outcome (confrontation intentions), and the additional mediators (personal efficacy, perpetrator malleability, affective and physical costs). Participants then indicated what race and gender they imagined the perpetrator to be in the vignettes and answered additional demographic questions about themselves. Finally, participants were thanked and debriefed.

Results

We excluded 27 participants who perceived that the actor in the vignettes was not European American yielding a sample of 525 participants (278 interpersonal condition, 247 structural condition). This final sample has 80% power to detect small effects (d=.25). See Table 4 for correlations among the dependent measures and Table 2 for descriptive statistics of the manipulation check measures. Consistent with our preregistered analyses, we conducted a series of t-tests to test the differences by experimental condition on the outcome and mediator measures. If the degrees of freedom for a t-test are not whole numbers, it indicates heterogeneity of variances (as tested with a Levene's test), so we reported the analysis correcting for this violated assumption (i.e., by utilizing un-pooled variances and correcting the degrees of freedom).

Manipulation check. Assessing the difference score (structural index-interpersonal index), a main effect of bias type emerged, $t(437.15)=8.78$, $p<.001$, $d=.78$; participants exposed to the structural vignettes (M=0.47, 95%CI [0.37, 0.56], SD=0.92) perceived that the vignettes reflected biased structures (vs. individuals with negative attitudes), more than those exposed to the interpersonal vignettes (M=-0.15, 95%CI [-0.25, -0.06], SD=0.65). Thus, the manipulation was successful (see Table 2).

Confrontation intentions. Consistent with Studies 1-2 and the preregistered hypothesis, participants reported that they would confront structural bias (M=3.49, 95%CI [3.34, 3.63], SD=1.12) more than interpersonal bias (M=3.24, 95%C [3.11, 3.37], SD=1.08), $t(523)=2.67$, $p=.008$, $d=.23$.

Perceived intentionality. Consistent with Studies 1-2 and the preregistered prediction, participants reported that perpetrating interpersonal bias (M=4.61, 95%CI [4.51, 4.71], SD=0.89)

was more intentional than perpetrating structural bias (M=4.15, 95%CI [4.04, 4.25], SD=0.80), t(523)=6.20, p<.001, d=.54.

Perceived harm. Aligned with predictions, participants reported structural bias vignettes (M=0.22, 95%CI [0.14, 0.28], SD=0.55) were more harmful than interpersonal bias vignettes (M=-0.15, 95%CI [-0.22, -0.08], SD=0.59), t(523)=7.29, p<.001, d=.69.

Personal efficacy. Contrary to Study 2 and the preregistered prediction, there was no difference in participants exposed to structural bias (M=2.84, 95%CI [2.68, 2.99], SD=1.27) and those exposed to interpersonal bias (M=2.78, 95%CI [2.63, 2.93], SD=1.25) in perceptions that they themselves would be effective confronters, t(523)=0.54, p=.591, d=.05.

Perpetrator malleability. Consistent with Study 2 and the preregistered prediction, participants perceived the perpetrator would listen more and be more likely to change if committing structural bias (M=3.03, 95%CI [2.92, 3.14], SD=0.84) than interpersonal bias (M=2.80, 95%CI [2.70, 2.91], SD=0.91), t(523)=2.93, p=.004, d=.26.

Perceived costs. Contrary to Study 2, no effect of bias type on perceived physical costs emerged, t(523)=1.60, p=.110, d=.14. However, means were in the predicted direction: participants perceived the perpetrator would act less aggressively if confronted for perpetrating structural bias (M=2.43, 95%CI [2.28, 2.57], SD=1.13) than interpersonal bias (M=2.59, 95%CI [2.45, 2.72], SD=1.17).

Consistent with Study 2, participants reported that structural bias (M=3.73, 95%CI [3.57, 3.90], SD=1.36) would be less awkward and uncomfortable to confront than interpersonal bias (M=4.13, 95%CI [3.97, 4.28], SD=1.23), t(497.83)=3.44, p=.001, d=.30.

Mediation. To assess what may drive participants' confrontation intentions for the different forms of bias (interpersonal vs. structural), we evaluated if perceptions of intentionality,

harm, personal efficacy, perpetrator malleability, physical costs, and affective costs mediated the relationship between type of bias and confrontation intentions (with 10,000 bootstrap samples, Hayes, 2017, model 4). Aligned with two of our three preregistered predictions regarding *why* participants would support confronting structural bias over interpersonal bias, people expressed a desire to confront someone perpetrating structural bias due to the perception that structural bias was more harmful and the perpetrator would change (perceived harm: 0.36, 95%CI [0.26, 0.47]; perpetrator malleability: 0.02, 95%CI [0.01, 0.05]; see Figure 6). Further, consistent with Studies 1-2, perceived intentionality mediated the effect of bias type on confrontation intentions [perceived intentionality: -0.04, 95%CI [-0.07, -0.02]; interpersonal bias was perceived as more intentional and more intentional behaviors were viewed as more necessary to confront. Contrary to Study 2 and the preregistered prediction, personal efficacy (0.03, 95%CI [-0.67, 0.13]) did not reliably mediate. Consistent with our preregistered predictions and Study 2, physical costs (-0.01, 95%CI [-0.03, 0.01]) and affective concerns (0.01, 95%CI [-0.01, 0.03]) also did not reliably mediate.

Overall, these results suggest that perceiving harm and that the perpetrator would change if confronted account for why European Americans supported confronting structural bias more than interpersonal bias. Additionally, while European Americans viewed interpersonal bias as more intentional and that more intentional behaviors should be confronted more, across studies, the total effect of bias type suggests that structural bias was viewed as more necessary to confront than interpersonal bias. This suggests that people may place greater weight on perceived harm and perpetrator malleability than perceived intentionality in their confrontation intentions regarding these types of bias.

Discussion

Study 3 provided additional evidence that European Americans support confronting a stranger's structurally-enacted bias more than interpersonal bias. The inclusion of additional vignettes with a wide variety of settings provided an important strength allowing us to better match vignettes across conditions and isolate the manipulation of bias type. Including these tightly-controlled vignettes, we again found that people reported greater willingness to confront structural bias than interpersonal bias.

Study 3 also tested potential explanations for this effect. Similar to research examining people's determinations of whether an act even qualifies as discrimination (e.g., Simon et al., 2019), intentionality perceptions also related to people's intentions to confront bias (consistent with Studies 1-2; see General Discussion). Participants' greater desire to confront structural bias, however, was better explained by perceptions that a perpetrator would be more likely to listen and change if confronted and that structural bias inflicts greater harm, than perceptions of intentionality. These results also compliment the collective action literature, which finds that feeling an action is efficacious and perceiving injustice will lead to action by advantaged group members (e.g., van Zomeren et al., 2008; Hornsey et al., 2018). Contrary to Study 2 and our preregistered predictions, participants did not indicate that they would personally be more effective in changing behavior if imagining confronting structural bias, compared to interpersonal bias. One possible explanation for this may involve the addition of new settings where the bias was perpetrated (e.g., in a store, at work) in Study 3. Future studies could directly assess how bias displayed in certain spaces may influence feelings of personal efficacy and confrontation intentions. Overall, using well-matched vignettes and drawing on relevant research on collective action, the results of Study 3 highlight important considerations (i.e., perceived

harm and perpetrator malleability) underlying European Americans' intentions to confront structural bias.

General Discussion

The present research assessed the conditions under which European Americans intend to confront different types of racial bias (interpersonal or structural) perpetrated by different actors (friends or strangers). Across three experiments, European Americans reported intentions to confront instances of structural bias more than instances of interpersonal bias, particularly if considering a stranger perpetrating the bias. Greater support for confronting structural racism was due to perceptions that structural racism is more harmful and that confronting will be more efficacious (i.e., the perpetrator will change). When exposed to interpersonal racial bias, European Americans reported greater desires to confront their friends than strangers. Overall, this work demonstrates that considering the various manifestations of bias—both looking at different types of bias and who perpetrates it—leads to a more complete understanding of what motivates confrontation decisions.

Expanding Confrontation Considerations: Perceived Harm, Confronter Identity, and Perpetrator Malleability

This research expands the confrontation literature by testing what processes underlie European Americans' intentions to confront people engaging in or contributing to biased structures and to confront people in their inner circle. Past work on recognizing discrimination indicates that advantaged group members, particularly European Americans, are more concerned with an actor's intent and less so on harm when determining if an act qualifies as discrimination (Simon et al., 2019). The present work, however, suggests that when evaluating whether to *confront* instances of structural bias, perceived harm looms larger than perceived intent. This

finding is aligned with the collective action and allyship literatures finding that advantaged group allies seek to engage in actions to address biased structures when recognizing the impact and pervasiveness of injustice (see Craig et al., 2020; Louis et al., 2019; Radke et al., 2020 for reviews). Taking our results together with this prior work, this suggests that perceived intent may drive when European Americans label an act as discrimination, but perceived harm may be crucial for driving support for confrontation.

Our work also contributes to the understanding of how potential confronters' identities may shape the processes driving confrontation intentions. Past research finds that targets of bias often undergo a cost-benefit analysis when deciding to confront interpersonal bias. Specifically, people are concerned with social backlash (e.g., being labeled as complainers, Good et al., 2012; Kaiser & Miller, 2004), their physical safety (e.g., Ayres et al., 2009), and the notion that the perpetrators are unwilling to change (Good et al., 2012; Rattan, 2019). Research assessing these costs primarily focus on *targets'* confrontation intentions. The present work examined how these factors may influence non-targeted (advantaged) group members' confrontation intentions and utilized ambiguously-biased scenarios. Our results suggest that physical safety and affective concerns may be less impactful for observers' intentions to confront (particularly structural bias or friend-perpetrators). Consistent with Rattan and Dweck (2010), however, our results indicate that confrontation intentions for observers are particularly influenced by perceptions that the perpetrator will listen and change if confronted (that the confrontation will be effective), which spurs support for confrontations of structural bias and friends.

The notion that people may confront friends and structural bias due to perceptions that the perpetrator would change and listen is also in line with the literature on bystander helping and empathy (e.g., Cameron et al., 2016, 2019). For example, efficacy concerns (is empathy

successful for achieving goals; Cameron et al., 2019) can deter engagement in empathy and people especially avoid empathy when interacting with strangers (Ferguson et al., 2020). Thus, the common concerns may underlie bystander helping and confrontation decisions. Future work could assess how these concerns may affect other actions meant to reduce bias (e.g., participating in protests).

Recognizing Structural Bias and Harm

In the vignettes utilized in our work, resulting disparities between racial groups were highlighted (e.g., pointing out that rezoning school districts would lead to more African American students attending lower funded schools). However, in everyday life it may be difficult to discern if an action will inflict harm before the behavior occurs. For example, people may not recognize the harm until the new school zones are already in place (if this harm is perceived at all). This possibility is of particular concern for structural bias. If not informed of the consequences, in situations with structural bias operating, European Americans may fail to detect bias and potential harms (Bonam et al., 2019). While our research shows that the recognition of harm from structural bias can facilitate support for confronting those enacting biased policies and structures, future research can evaluate strategies to increase recognition of potential harm from structural bias in the first place (Bonilla-Silva, 2015; Trawalter et al., 2019).

Importantly, to cleanly compare interpersonal and structural bias, we attempted to keep constant an individual perpetrator in our vignettes. As a result, we empirically examined one form in which structural bias may occur and be confronted. However, structural bias may already exist in the world and multiple actors may support it and allow it to continue (Bonilla-Silva, 1997). Future research can assess how people seek to address other instantiations of structural

bias (e.g., biased structures created and maintained by multiple powerful individuals or upheld throughout history by society-at-large) to capture the broader construct of structural bias.

Once people are presented with structural bias, work in system justification and hierarchy maintenance (e.g., Jost, Banaji, & Nosek, 2004; Knowles et al., 2014) may predict that individuals would seek to justify the inequitable system, given that people (and particularly those advantaged by a system) often seek to justify the status quo and defend existing status hierarchies. However, our data highlight a situation in which European Americans recognized the harms resulting from an unjust system and this harm perception led to support for intervention. Perhaps participants recognized the harm from structural bias in our studies because we did not label the instances as explicitly unfair or unjust (as system threat manipulations may do; see Kay & Friesen, 2011; Kay, Jost, & Young, 2005) or because our vignettes were narrow in scope (presenting bias at the local-level vs. the national-level), thus allowing people to recognize the harms in the system for themselves or allowing people to focus on local-level harms as opposed to national-level harms. Future work should delineate the circumstances in which encountering information about structural or systemic social problems may enhance or mitigate system justification and hierarchy maintenance processes.

Limitations and Future Directions

The primary limitation to the current work is that people imagined a situation in which they could confront (or not). Past work indicates that people may be inaccurate in forecasting their actual confrontation behavior (see Kawakami et al., 2019). It's possible that people may forecast that they *would* confront their friends more than strangers and structural bias over interpersonal bias, as the present work reveals, but fail to confront in the moment. Importantly, the studies assessing the gap between forecasting and actual confrontation (see Kawakami et al.,

2019) focused on witnessing blatant acts of interpersonal bias. It's not clear whether or not people accurately forecast confrontation of subtly-expressed structural bias and friend perpetrators. Future studies that allow for actual confrontation of different bias severity vignettes (i.e., blatant or subtle bias) and manifestations of bias (i.e., structural or interpersonal) can further clarify what influences behavioral confrontation.

Importantly, work on intergroup contact shows the impact of imagined scenarios. Imagining positive contact with outgroup members may facilitate positive intergroup relations, influencing attitudes, stereotypes, and engagement in future contact (Crisp et al., 2009; Crisp & Turner, 2012; Miles & Crisp, 2013). If confrontation works through similar processes, imagining confronting and considering what leads to confronting may influence engagement in actual confrontations when witnessing bias in the future. Future studies can capture the influence of imagined confrontations on future confronting by having participants imagine confronting (vs. remaining silent towards) a friend or stranger who acted biased and measuring decisions to confront an actual future incident. If an imagined confrontation leads people to engage in future confrontation, this would aid in the understanding of the impact of imagined behavior (e.g., Weidman et al., 2019).

In addition, the current experiments rely exclusively on vignettes depicting instances of *racial* bias with European American perpetrators and US participants. It is possible that our results would replicate for different dimensions of inequality (e.g., sexism, homophobia, ableism) and other nations; however, this is speculative, given that our studies only focused on racial bias in the US. Further, seeing as European American friendship networks often over-represent European Americans than members of other groups (Ingraham, 2014), we focused on decisions to confront European American perpetrators. It's an open question how the decision-

making process unfolds if considering confrontation of perpetrators with different identities (e.g., someone who identifies as Asian American or a European American woman who highlights her gender identity). Future studies should assess how the identity of perpetrators may impact confrontations of different manifestations of bias.

Conclusion

Across three experiments, European Americans reported support for confronting instances of structural racial bias more than interpersonal racial bias and that a friend should be confronted more than a stranger for biased behavior. Considering the types of bias and who perpetrates helps elucidate the different obstacles to confrontation and may aid in the development of effective interventions. Educating people on how structural biases may manifest (e.g., people instituting policies or following norms) and the potential harms of structural bias may be especially potent in eliciting ally action (e.g., Adams et al., 2008; Craig et al., 2020; Unzueta & Lowery, 2008). Within the past few years in the U.S., there has been an uptick in federal rules that negatively impact marginalized communities (Green, 2019). These acts of structural bias (for example, the confrontation regarding structural bias in Flint, Michigan, described in the introduction), elucidate how vital and timely it is to understand what motivates people to address the many ways in which bias may emerge.

Footnotes

[1]Participants read four interpersonal and four structural vignettes. To ensure the vignettes were similar within bias conditions, we conducted analyses comparing the other vignettes within each bias condition on the main dependent measures. We also conducted these analyses because our manipulation check items did not cleanly capture participants' perception that these vignettes reflected acts of interpersonal or structural bias (we include a better manipulation check in Studies 2-3). We excluded one interpersonal and one structural vignette after the analyses revealed that the vignettes differed significantly from the other vignettes. If including all original vignettes in the main analyses, main effects of bias type on intentionality ratings are marginal and main effects of bias type on confrontation intentions are trending. See supplemental materials for these analyses and the full analysis including the additional vignettes.

[2]See supplement for manipulation check analyses that use indices of interpersonal perceptions and structural perceptions as a within-subjects variable for Studies 1-3. In Study 1, participants also rated how much each vignette reflects someone with negative attitudes towards racial minorities as an additional manipulation check. However, this item only weakly correlated with the main manipulation check (the interpersonal perceptions index; r=.232, p=.001). Thus, we did not include this item in the manipulation check analyses.

[3]Participants completed exploratory items assessing how much each scenario could be attributed to discrimination and perceptions that general racism is intentional.

[4]The number of participants who were excluded from the analysis did not differ by condition, $\chi(3)=4.42$, $p>.220$.

[5]To test if perceived intentionality drove participants' support for confronting interpersonal bias, replicating Study 1, we again tested the mediating effect of perceived

intentionality on the relationship between bias type and confrontation intentions (the effect replicated). See supplemental materials for this analysis.

[6]The difference score for the manipulation check was used to remain consistent with Studies 1-2. The preregistered report indicated that we would look separately at how much the participants perceived the vignettes to depict either interpersonal or structural bias. Aligned with the preregistration, participants who read the interpersonal vignettes rated them as more of a reflection of someone with negative attitudes towards racial minorities compared to those who read the structural vignettes. Participants who read the structural vignettes tended (not significantly) to rate them as more of a reflection of common practices and norms than participants who read the interpersonal vignettes. See supplemental materials for these analyses.

[7]In the preregistration, we stated that we would analyze harm and pervasiveness separately. However, the high correlation between the two scales (r=.78, p<.001) suggests that participants did not differentiate these constructs, thus we combined them into one perceived harm index in the main text (see the supplement for the separated analyses, which mirror the results reported in the main text).

Open Practices

The experiments in this article earned an Open Data and Preregistered badge for transparent practices. Data for Studies 1-3 are available at https://osf.io/gvchf2/. The preregistration document for Study 3 can be found here: http://aspredicted.org/blind.php?x=fm6gx6.

References

Adams, G., Edkins, V., Lacka, D., Pickett, K. M., & Cheryan, S. (2008). Teaching about racism: Pernicious implications of the standard portrayal. *Basic and Applied Social Psychology*, *30*(4), 349–361.

Andersen, S. M., & Chen, S. (2002). The relational self: an interpersonal social-cognitive theory. *Psychological Review*, *109*(4), 619.

Apfelbaum, E. P., Grunberg, R., Halevy, N., & Kang, S. (2017). From ignorance to intolerance: Perceived intentionality of racial discrimination shapes preferences for colorblindness versus multiculturalism. *Journal of Experimental Social Psychology, 69*, 86-101.

Aron, A., & Aron, E. N. (1986). *Love and the expansion of self: Understanding attraction and satisfaction*. Hemisphere Publishing Corp/Harper & Row Publishers.

Ayres, M. M., Friedman, C. K., & Leaper, C. (2009). Individual and situational factors related to young women's likelihood of confronting sexism in their everyday lives. *Sex Roles*, *61*(7-8), 449-460.

Ashburn-Nardo, L., & Abdul Karim, M. F. (2019). *The CPR model*. *Confronting Prejudice and Discrimination*. In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Elsevier.

Ashburn-Nardo, L., Morris, K. A., & Goodwin, S. A. (2008). Responses (CPR) Model : Applying CPR in Organizations. *Academy of Management Learning & Education*, *7*(3), 332–342.

Blodorn, A., O'Brien, L. T., & Kordys, J. (2012). Responding to sex-based discrimination: Gender differences in perceived discrimination and implications for legal decision making. *Group Processes and Intergroup Relations*, *15*(3), 409–424.

Booker, B. (2021). *Ex-Michigan Gov. Rick Snyder And 8 Others Criminally Charged In Flint Water Crisis.* NPR.org. https://www.npr.org/2021/01/14/956924155/ex-michigan-gov-rick-snyder-and-8-others-criminally-charged-in-flint-water-crisi

Bonam, C. M., Nair Das, V., Coleman, B. R., & Salter, P. (2019). Ignoring history, denying racism: Mounting evidence for the Marley hypothesis and epistemologies of ignorance. *Social Psychological and Personality Science, 10*(2), 257-265.

Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American Sociological Review*, 465-480.

Bonilla-Silva, E. (2015). The structure of racism in color-blind, "Post-racial" America. *American Behavioral Scientist*, *59*(11), 1358–1376.

Cameron, C. D., Harris, L. T., & Payne, B. K. (2016). The emotional cost of humanity: Anticipated exhaustion motivates dehumanization of stigmatized targets. *Social Psychological and Personality Science, 7*(2), 105–112.

Cameron, C. D., Hutcherson, C. A., Ferguson, A. M., Scheffer, J. A., Hadjiandreou, E., & Inzlicht, M. (2019). Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General, 148*(6), 962–976.

Campanhã, C., Minati, L., Fregni, F., & Boggio, P. S. (2011). Responding to unfair offers made by a friend: neuroelectrical activity changes in the anterior medial prefrontal cortex. *Journal of Neuroscience*, *31*(43), 15569-15574.

Chaney, K. E., & Sanchez, D. T. (2018). The Endurance of Interpersonal Confrontations as a Prejudice Reduction Strategy. *Personality and Social Psychology Bulletin*, *44*(3), 418–429.

Chaney, K. E., Young, D. M., & Sanchez, D. T. (2015). Confrontation's health outcomes and promotion of egalitarianism (C-HOPE) framework. *Translational Issues in Psychological Science, 1*(4), 363.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1-8.

Corning, A., & Bucchianeri, M. (2010). Perceiving racism in ambiguous situations: Who relies on easy-to-use information? *Journal of Social Psychology*, *150*(3), 258–277.

Craig, M. A., Badaan, V., & Brown, R. M. (2020). Acting for whom, against what? Group membership and multiple paths to engagement in social change. *Current Opinion in Psychology,* 35, 41-48.

Craig, M. A., Brown, R. M., & Rucker, J. M. (unpublished manuscript). Structural solidarity: Lay theories of group disadvantage and coalitional attitudes among stigmatized groups.

Crisp, R. J., Stathi, S., Turner, R. N., & Husnu, S. (2009). Imagined intergroup contact: Theory, paradigm and practice. *Social and Personality Psychology Compass, 3*(1), 1-18.

Crisp, R. J., & Turner, R. N. (2012). The Imagined Contact Hypothesis. A*dvances in Experimental Social Psychology, 46,* 125-182 .

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.

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.

Dickter, C. L., & Newton, V. A. (2013). To confront or not to confront: Non-targets' evaluations of and responses to racist comments. *Journal of Applied Social Psychology*, *43*, 262-275.

Dietze, P., & Craig, M. A. (2020). Framing economic inequality and policy as group disadvantages (vs. group advantages) spurs support for action. *Nature Human Behaviour.*

Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations, 15*(4), 487-502.

Ferguson, A. M., Cameron, D., & Inzlicht, M. (2020). Motivational effects on empathic choices. https://doi.org/10.31234/osf.io/s7qph

Fryberg, S. A., & Eason, A. E. (2017). Making the invisible visible: Acts of commission and omission. *Current Directions in Psychological Science, 26(*6), 554-559.

Green, E. (2019). *How Trump is Reversing Obama's Nondiscrimination Legacy.* The Atlantic. https://www.theatlantic.com/politics/archive/2019/08/trump-lgbtq-rules/596116/

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

Guerin, B. (2005). Combating everyday racial discrimination without assuming racists or racism: New intervention ideas from a contextual analysis. *Behavior and Social Issues, 14*, 46–70.

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

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Heiphetz, L. & Craig, M. A. (2020). Links between Dehumanization and Perceptions of Immoral Intergroup Behavior. *Oxford Studies in Experimental Philosophy.*

Hornsey, M. J., Blackwood, L., Louis, W., Fielding, K., Mavor, K., Morton, T., ... & White, K. M. (2006). Why do people engage in collective action? Revisiting the role of perceived effectiveness. *Journal of Applied Social Psychology, 36*(7), 1701-1722.

Horowitz, J. M., Brown, A., & Cox K. (2019). *How Americans See the State of Race Relations.* Pew Research Center. https://www.pewsocialtrends.org/2019/04/09/how-americans-see-the-state-of-race-relations/#across-racial-and-ethnic-groups-similar-shares-say-they-hear-racist-or-racially-insensitive-comments-from-friends-or-family

Iyer, A., & Ryan, M. K. (2009). Why do men and women challenge gender discrimination in the workplace? The role of group status and in-group identification in predicting pathways to collective action. *Journal of Social Issues, 65*(4), 791-814.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology, 25*(6), 881-919.

Kaiser, C. R., & Miller, C. T. (2004). A stress and coping perspective on confronting sexism. *Psychology of Women Quarterly, 28*(2), 168-178.

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and behavioral responses to racism. *Science*, *323*(1), 276–278.

Kawakami, K., Karmali, F., Vaccarino, E. (2019). Confronting intergroup bias: Predicted and actual responses to racism and sexism. In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Elsevier.

Kay, A. C., & Friesen, J. (2011). On social stability and social change: Understanding when system justification does and does not occur. *Current Directions in Psychological Science, 20*(6), 360-364.

Kay, A. C., Jost, J. T., & Young, S. (2005). Victim derogation and victim enhancement as alternate routes to system justification. *Psychological Science, 16*(3), 240-246.

Knowles, E. D., Lowery, B. S., Chow, R. M., & Unzueta, M. M. (2014). Deny, distance, or dismantle? How white Americans manage a privileged identity. *Perspectives on Psychological Science, 9*(6), 594-609.

Louis, W. R., Thomas, E., Chapman, C. M., Achia, T., Wibisono, S., Mirnajafi, Z., & Droogendyk, L. (2019). Emerging research on intergroup prosociality: Group members' charitable giving, positive contact, allyship, and solidarity with others. *Social and Personality Psychology Compass, 13*(3), e12436.

Martin, R. (2021). *Flint Pediatrician Says Charges Against Ex-Governor Help 'Wounds Finally Close'.* NPR.org. https://www.npr.org/2021/01/14/956781868/flint-pediatrician-says-charges-against-ex-governor-help-wounds-finally-close

Marques, J. M., Yzerbyt, V. Y., & Leyens, J. P. (1988). The “black sheep effect”: Extremity of judgments towards ingroup members as a function of group identification. *European Journal of Social Psychology, 18*(1), 1-16.

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations, 17*(1), 3-26.

National Healthcare Quality and Disparities Reports. Content last reviewed April 2019. Agency for Healthcare Research and Quality, Rockville, MD.

Nelson, J. C., Adams, G., & Salter, P. S. (2013). The Marley Hypothesis: Denial of Racism

Reflects Ignorance of History. *Psychological Science*, *24*(2), 213–218.

O'Brien, L. T., Blodorn, A., Alsbrooks, A., Dube, R., Adams, G., & Nelson, J. C. (2009). Understanding White Americans' perceptions of racism in hurricane Katrina-related events. *Group Processes & Intergroup Relations*, *12*(4), 431-444.

Otten, S. (2009). Social categorization, intergroup emotions and aggressive interactions. *Intergroup Relations: The Role of Motivation and Emotion,* 162-181.

Ozier, E. M., Taylor, V. J., & Murphy, M. C. (2019). The cognitive effects of experiencing and observing subtle racial discrimination. *Journal of Social Issues,* 1-29.

Pinto, I. R., Marques, J. M., Levine, J. M., & Abrams, D. (2010). Membership status and subjective group dynamics: Who Triggers the Black Sheep Effect? *Journal of Personality and Social Psychology*, *99*(1), 107.

Radke, H. R., Kutlaca, M., Siem, B., Wright, S. C., & Becker, J. C. (2020). Beyond allyship: Motivations for advantaged group members to engage in action for disadvantaged groups. *Personality and Social Psychology Review*, *24*(4), 291-315.

Rasinski, H. M., & Czopp, A. M. (2010). The Effect of Target Status on Witnesses' Reactions to Confrontations of Bias. *Basic and Applied Social Psychology*, *32*(1), 8–16.

Rattan, A. (2019). How lay theories (or mindsets) shape the confrontation of prejudice. In R. K. Mallett, & M. J. Monteith (Eds.), *Confronting prejudice and discrimination: The science of changing minds and behaviors*. Elsevier.

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

Salter, P. S., Adams, G., & Perez, M. J. (2018). Racism in the structure of everyday worlds: A cultural-psychological perspective. *Current Directions in Psychological Science, 27(*3), 150-155.

Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group members. *Evolution and Human Behavior, 25*(6), 379-393.

Simon, S., Moss, A. J., & O'Brien, L. T. (2019). Pick your perspective: Racial group membership and judgments of intent, harm, and discrimination. *Group Processes and Intergroup Relations*, *22*(2), 215–232.

Southern Poverty Law Center. (2015, January). *Speak up: Responding to everyday bigotry.* SPLC. https://www.splcenter.org/20150125/speak-responding-everyday-bigotry

Swim, J. K., Scott, E. D., Sechrist, G. B., Campbell, B., & Stangor, C. (2003). The role of intent and harm in judgments of prejudice and discrimination. *Journal of Personality and Social Psychology, 84*, 944–959.

Trawalter, S., Bart-Plange, D.-J., & Hoffman, K. M. (2019). A socioecological psychology of racism: Making structures and history more visible. *Current Opinion in Psychology, 32.*

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In *Advances in Experimental Social Psychology*, 21, 181-227.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worschel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Brooks/Cole.

Unzueta, M. M., & Lowery, B. S. (2008). Defining racism safely: The role of self-image maintenance on white Americans' conceptions of racism. *Journal of Experimental Social*

Psychology, *44*(6), 1491–1497.

Van Zomeren, M. (2013). Four core social-psychological motivations to undertake collective action. *Social and Personality Psychology Compass, 7*(6), 378-388.

Van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. *Psychological Bulletin, 134*(4), 504.

Weidman, A. C., Sowden, W. J., Berg, M. K., & Kross, E. (2019). Punish or protect? How close relationships shape responses to moral violations. *Personality and Social Psychology Bulletin*, *16*(5), 693-708.

Williams, D. R., & Mohammed, S. A. (2009). Discrimination and racial disparities in health: evidence and needed research. *Journal of Behavioral Medicine, 3*2(1), 20-47.

Wang, L., Zheng, J., Meng, L., Lu, Q., & Ma, Q. (2016). Ingroup favoritism or the black sheep effect: Perceived intentions modulate subjective responses to aggressive interactions. *Neuroscience Research*, *108*, 46–54.

Appendix

Table 1A
Vignettes across conditions in Studies 1-2:

Bias Condition			
Interpersonal	Yesterday Jamie was on the downtown bus. After a few stops, a black family boarded and sat down near Jamie. Shortly after the family sat down, (s)he got up, walked down the aisle, and held a handrail. Jamie did not get off at the next stop.	Jamie works for Uber. One day, Jamie was driving around town waiting to pick up passengers. As he was stopped at a red light, he noticed a black man walking down the sidewalk on the same side of the street. The man was dressed in layers and appeared to be carrying a bundle of newspapers. Upon seeing the man, Jamie locked the car doors.	A weeks ago, Jamie dined in a restaurant that was very busy. At another table, he/she saw a black couple sit for 30 minutes before getting served. Three couples were served before them, two of which had come into the restaurant after the black couple. Over the noise, Jamie overheard their waiter, who was white, tell the hostess that another server would need to cover that table because he was too busy. Jamie left the waiter a big tip and wrote a compliment card about the great job he had done.
Structural	Jamie went to the local town hall. During the session, the council proposed a motion that would divert money from local transportation initiatives to predominantly-black communities exposed to environmental hazards (e.g., lead paint and diesel fumes). During the meeting, Jamie suggested that there was nothing we could do about this problem and that the money would be more useful for education.	The school board in your community recently proposed a change to the school district zones due to overcrowding at some schools. Before the order passed, a community member noted that most of the black students were redistricted to lower funded schools, while white students were placed in more funded schools. Despite this concern, most of the town, including Jamie, voted in favor of the law.	Two candidates were recently hired to a full-time job at a small store. Jamie works in the company's HR department, where he stumbled on paperwork showing that the new hire who is black received a lower base salary than the new hire who is white. Jamie decided not to tell anyone what he had seen.

Table 2A

Vignette matching across conditions in Study 3:
considering setting, commission/omission, if target(s) recognized their being affected

Qualities	**Setting**					
	Bus	Uber	Town Hall	Small Store	Realty Office	Hair
Interpersonal	Yesterday Jamie was on the downtown bus. After a few stops, a black family boarded and sat down near Jamie. Shortly after the family sat down, (s)he got up, walked down the aisle, and held a handrail. Jamie did not get off at the next stop. **Studies 1-3**	Jamie works for Uber. One day, Jamie was driving around town waiting to pick up passengers. As he was stopped at a red light, he noticed a black man walking down the sidewalk on the same side of the street. The man was dressed in layers and appeared to be carrying a bundle of newspapers. Upon seeing the man, Jamie locked the car doors. **Studies 1-3**	Jamie is in charge of moderating questions at the local town hall. Yesterday, the council proposed a motion to assess where to allocate new funds, either to transportation or to education. During the session, he/she calls on the white community members more often than the black community members. **Study 3**	Jamie recently started a new full-time job at a small store. One Saturday, he/she noticed a black person entered the business and begin looking at some of the items on display. Immediately, Jamie approached the person and greeted him by asking, "Are you here to buy something today?" **Study 3**	Jamie is a realtor and works in a nice part of town. Two families came into the realtor office to search for a realtor, one black family and one white family. Jamie went straight to ask if the white family needed help. **Study 3**	Jamie works at a local company. He/she was talking with his/her white coworker. When one of their black coworkers walked by, Jamie made a joke about the black coworkers who wore their hair natural. His/her white coworker laughed, and they went back to their desk. **Study 3**
Act of Commission (action) or Omission (non-action)	Commission	Commission*	Commission	Commission	Commission	Commission
Target(s) did or did not notice the bias	Noticed	Did not notice	Noticed	Noticed*	Noticed	Did not notice*
Structural	Jamie works for a bus company. The new route he/she created to save time in the morning ends	Jamie works for Uber and is in charge of granting raises to his employees. Jamie approved a system of	Jamie went to the local town hall. During the session, the council proposed a motion that would	Two candidates were recently hired to a full-time job at a small store. Jamie works in the company's HR	Jamie is a realtor in charge of finding a place for a new strip mall downtown in	Jamie works at a local company. He/she is in charge of creating and enforcing company

	up creating new stops in the predominantly white community and taking away stops in the predominantly black community. **Study 3**	granting raises that makes bonuses proportion to base pay. (S)he recently met with two potential candidates for a raise, a white employee and a black employee. Jamie granted the white employee with a larger raise, because the white employee started with a larger base pay. **Study 3**	divert money from local transportation initiatives to predominantly-black communities exposed to environmental hazards (e.g., lead paint and diesel fumes). During the meeting, Jamie suggested that there was nothing we could do about this problem and that the money would be more useful for education. **Studies 1-3**	department, where he stumbled on paperwork showing that the new hire who is black received a lower base salary than the new hire who is white. Jamie decided not to tell anyone what he had seen. **Studies 1-3**	his/her city. When he/she finally chooses a location, his/her plan results in the displacement of a large number of black people from their home and community. **Study 3**	dress codes. He/she had been getting complaints about the casual nature of the office and decided to make a new policy where employees must keep their hair neat. In turn, black coworkers who kept their hair natural no longer met the professional dress code. **Study 3**
Act of Commission (action) or Omission (non-action)	Commission	Omission*	Commission	Commission	Commission	Commission
Target(s) did or did not notice the bias	Noticed	Did not notice	Noticed	Did not notice*	Noticed	Noticed*

Note. Stars (*) indicate significant differences at $p<.05$ across condition matched pairs in a pilot survey of the vignettes.

Tables and Figures

Table 1

Correlations between dependent measures in Study 1 and Study 2

	Study 1		*Study 2 (bias type condition is within-subjects)*									
			Interpersonal bias condition					Structural bias condition				
	1	2	1	2	3	4	5	1	2	3	4	5
1. Confrontation Intentions	--		--					--				
2. Perceived Intentionality	--	.23***	--	.21***				--	.22***			
3. Personal Efficacy	--	--	.69***	.02	--			.69***	-.01	--		
4. Perpetrator Malleability	--	--	.59***	-.12*	.63***	--		.56***	-.12*	.62***	--	
5. Perceived Affective Cost	--	--	-.12†	.15**	-.41***	-.18**	--	-.14*	.07	-.39***	-.19***	--
6. Perceived Physical Cost	--	--	.14*	.14*	-.16**	-.05	.51***	.11†	.12*	-.16**	-.08	.49***

Notes. $^{\dagger}p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

Table 2

Descriptive statistics for manipulation check items by bias type experimental condition.

	Interpersonal bias condition		Structural bias condition	
	M (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Study 1				
Interpersonal Perceptions	5.08_a (1.24)	[4.85, 5.33]	4.71_b (1.21)	[4.45, 4.96]
Structural Perceptions	4.64_b (1.57)	[4.33, 4.91]	4.91_b (1.42)	[4.57, 5.20]
Study 2				
Interpersonal Perceptions	4.05_a (1.36)	[3.90, 4.20]	3.83_b (1.37)	[3.68, 3.98]
Structural Perceptions	3.78_b (1.45)	[3.62, 3.95]	4.36_c (1.43)	[4.02, 4.52]
Study 3				
Interpersonal Perceptions	4.04_a (1.04)	[3.90, 4.17]	3.56_b (1.22)	[3.42, 3.70]
Structural Perceptions	3.88_c (1.21)	[3.74, 4.03]	4.03_c (1.31)	[3.87, 4.18]

Notes. Means with different subscripts indicate significant differences at $p<.05$.

Table 3

Studies 1-2: Descriptive statistics by experimental conditions.

	Interpersonal bias condition				Structural bias condition			
	Stranger condition		Friend condition		Stranger condition		Friend condition	
	M (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]	*M* (*SD*)	[95% CI]
Study 1								
Confrontation Intentions	3.09_a (1.51)	[2.70, 3.48]	3.87_b (1.46)	[3.52, 4.22]	3.92_b (1.23)	[3.57, 4.27]	4.12_b (1.06)	[3.69, 4.56]
Perceived Intentionality	4.69_a (1.22)	[4.37, 5.00]	4.80_a (1.09)	[4.52, 5.08]	4.46_a (1.12)	[4.18, 4.74]	4.24_b (0.85)	[3.88, 4.60]
Study 2								
Confrontation Intentions	2.73_a (1.23)	[2.59, 2.86]	3.28_b (1.49)	[3.14, 3.42]	3.40_c (1.31)	[3.26, 3.54]	3.65_c (1.52)	[3.51, 3.79]
Perceived Intentionality	5.03_a (0.83)	[4.89, 5.17]	4.77_b (1.00)	[4.62, 4.92]	4.77_b (0.82)	[4.62, 4.92]	4.43_c (1.06)	[4.28, 4.58]
Personal Efficacy	2.29_a (1.22)	[2.07, 2.52]	3.70_b (1.63)	[3.47, 3.93]	2.59_c (1.25)	[2.36, 2.81]	3.70_b (1.59)	[3.47, 3.92]
Perpetrator Malleability	2.05_a (0.91)	[1.87, 2.22]	3.13_b (1.27)	[2.95, 3.30]	2.59_c (0.91)	[2.40, 2.77]	3.46_d (1.34)	[3.29, 3.64]
Perceived Affective Cost	4.17_a (1.51)	[3.93, 4.42]	3.09_b (1.62)	[2.83, 3.33]	3.94_c (1.47)	[3.70, 4.19]	3.04_b (1.60)	[2.80, 3.29]
Perceived Physical Cost	2.72_a (1.35)	[2.54, 2.91]	1.79_b (0.99)	[1.60, 1.98]	2.59_c (1.22)	[2.41, 2.77]	1.76_b (1.04)	[1.58, 1.94]

Notes. Means with different subscripts indicate significant differences at $p<.05$.

Table 4

Zero-order Correlations between measures in Study 3

	1	2	3	4	5	6
1. Confrontation Intention	--					
2. Perceived Intentionality	.32***	--				
3. Perceived Harm	.79***	.31***	--			
4. Personal Efficacy	.77***	.16***	.40***	--		
5. Perpetrator Malleability	.49***	-13**	.36***	.45***	--	
6. Perceived Affective Cost	-.03	.04	.15**	-.28***	.05	--
7. Perceived Physical Cost	.34***	.14***	.35***	.11**	.24***	.45***

Notes. $^{\dagger}p < .10$, $*p < .05$, $**p < .01$, $***p < .001$.

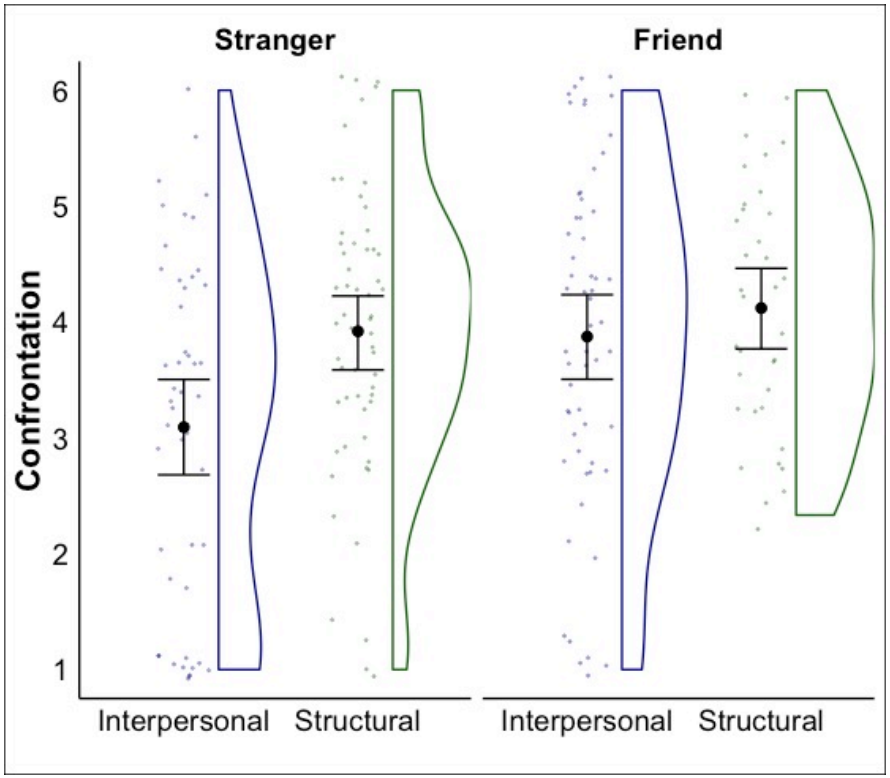

Figure 1: Raincloud plot with jittered data for participants' confrontation intentions for different types of bias (interpersonal condition, structural condition) depending on the closeness of the perpetrator (stranger, friend) in Study 1. Error bars represent 95% confidence intervals around the mean (black dots). Blue and green dots indicate individual participant scores. The "cloud" areas reflect the data distributions.

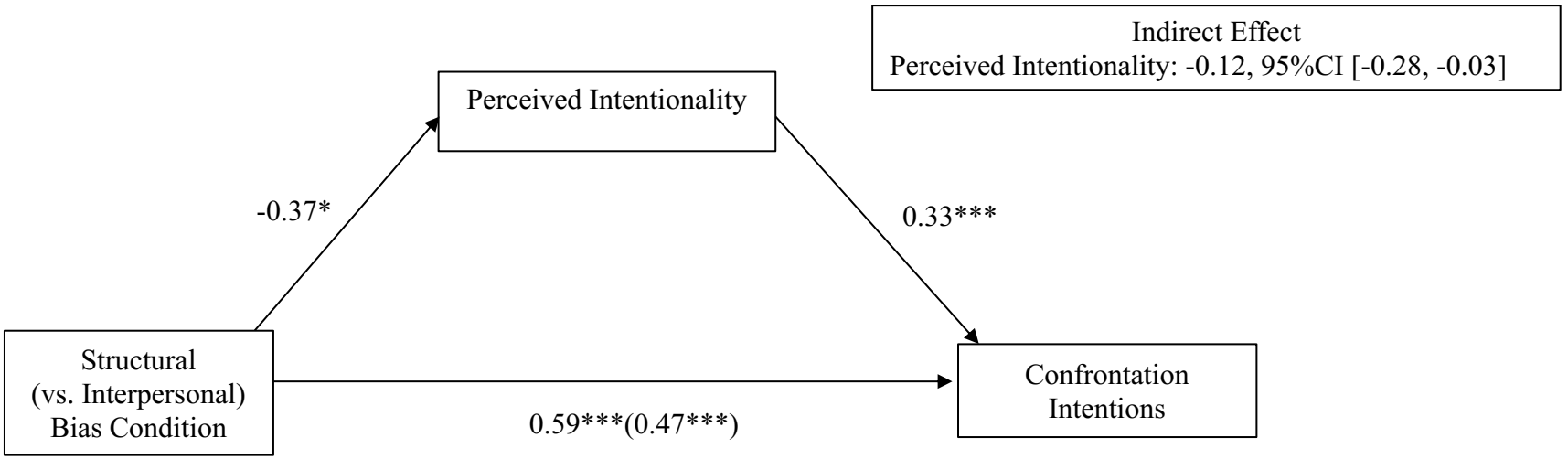

Figure 2: Study 1 process model of the indirect effect of bias condition (structural condition vs. interpersonal condition) on confrontation intention via perceived intentionality (with 10,000 bootstrap samples). The values in parentheses represent the total effects prior to the inclusion of the mediator. *Notes.* *p<.05, **p < .01, ***p < .001.

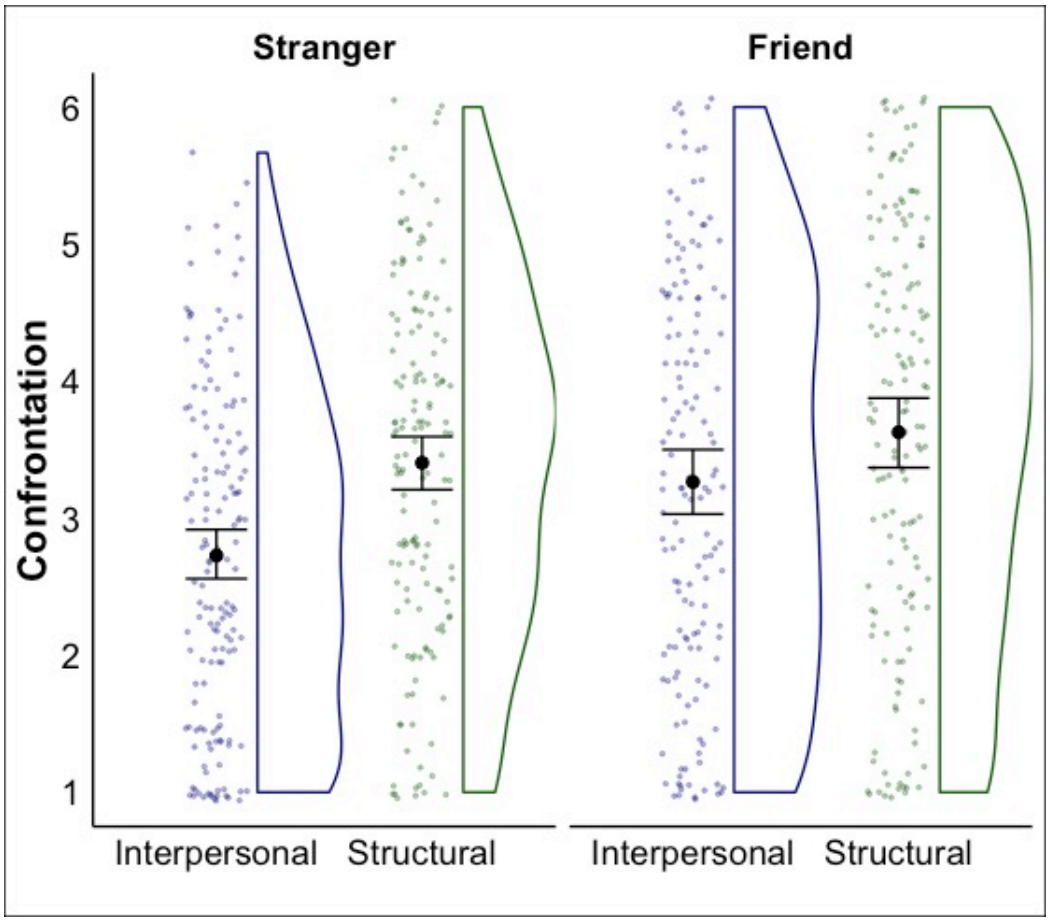

Figure 3: Raincloud plot with jittered data for participants' confrontation intentions for different types of bias (interpersonal condition, structural condition) depending on the closeness of the perpetrator (stranger, friend) in Study 2. Error bars represent 95% confidence intervals around the mean (black dots). Blue and green dots indicate individual participant scores. The "cloud" areas reflect the data distributions.

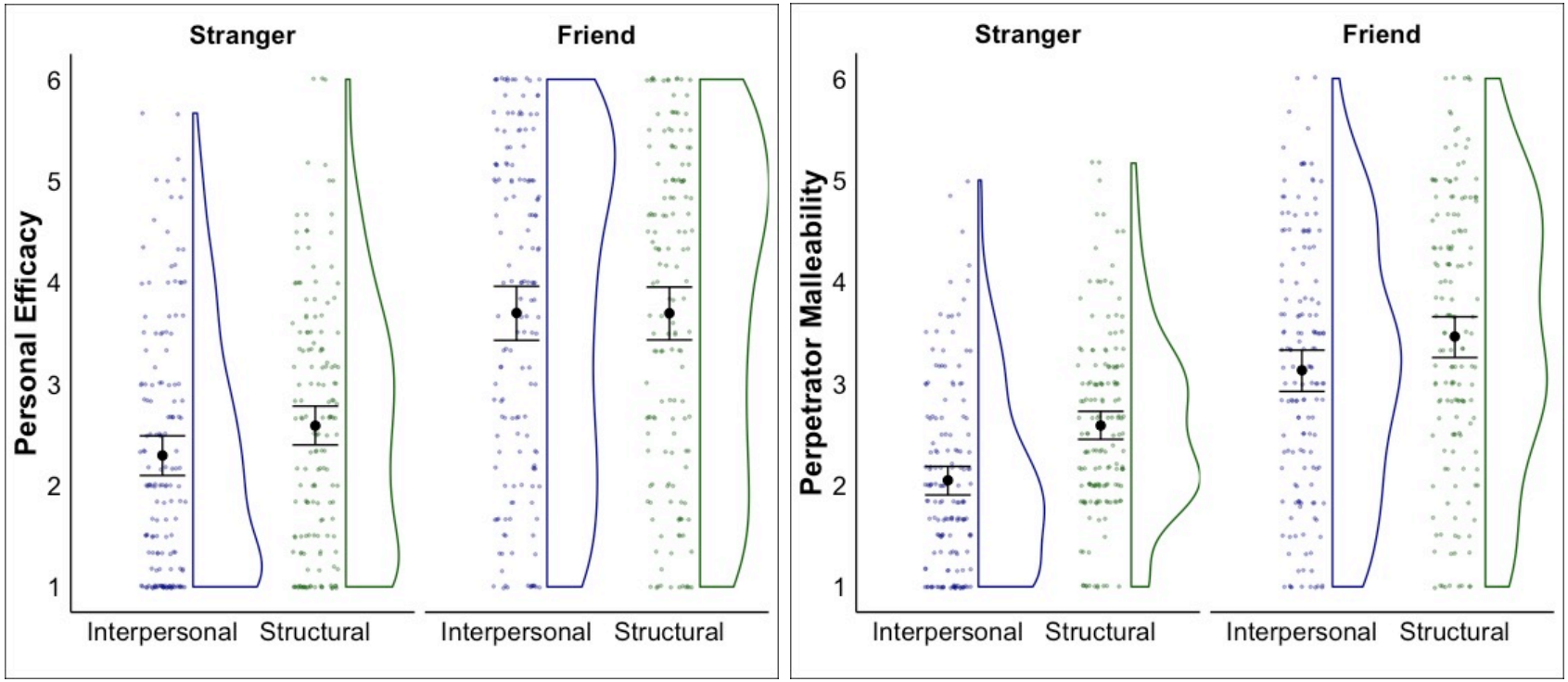

Figure 4: Raincloud plot with jittered data for personal efficacy ratings and perceived perpetrator malleability ratings for different types of bias (interpersonal condition, structural condition) depending on the closeness of the perpetrator (stranger, friend) in Study 2. Error bars represent 95% confidence intervals around the mean (black dots). Blue and green dots indicate individual participant scores. The "cloud" areas reflect the data distributions.

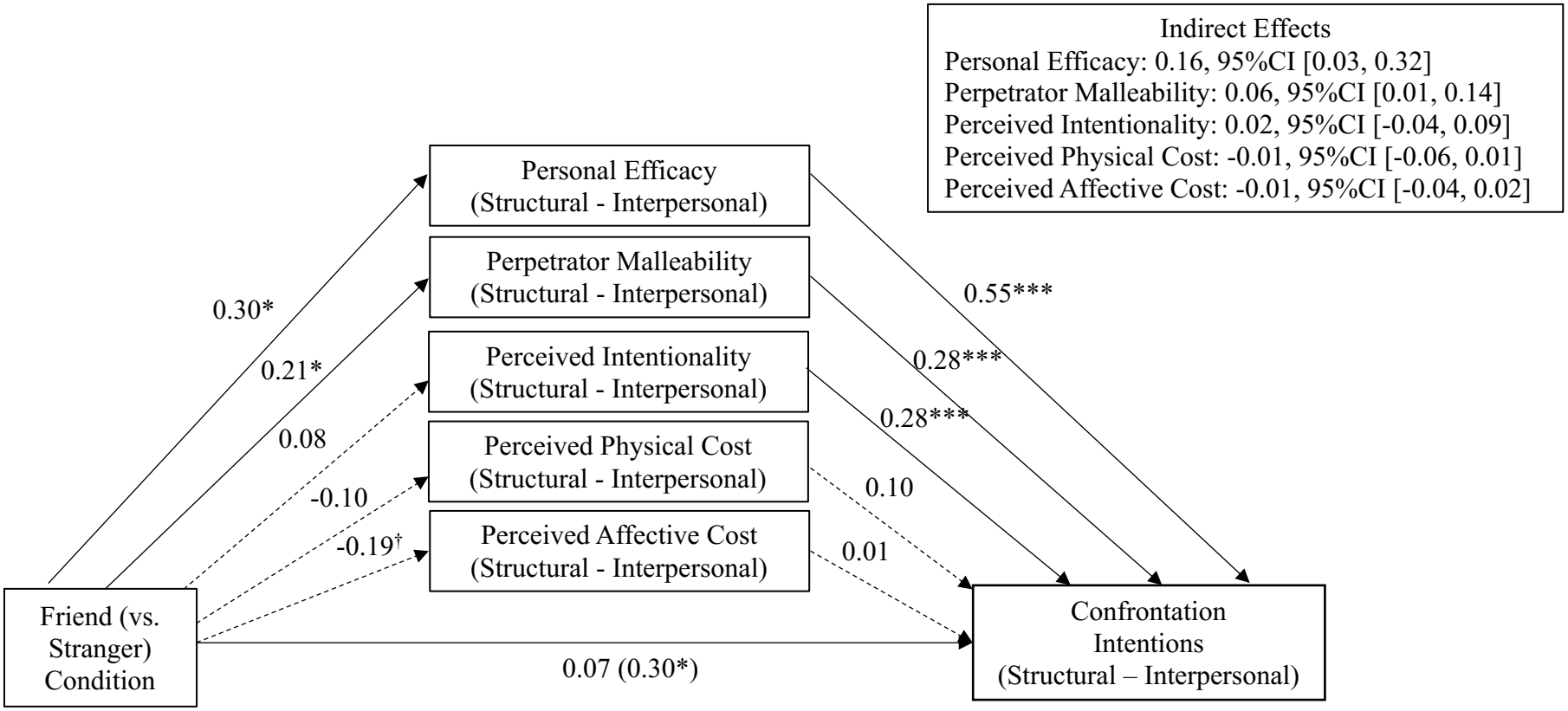

Figure 5: Testing for statistical mediation of the interactive effect between perpetrator closeness (between-subjects factor; friend vs. stranger) and bias type (within-subjects factor; structural condition vs. interpersonal condition) on confrontation intentions in Study 2 via personal efficacy, perceived perpetrator malleability, perceived intentionality, perceived physical costs, and perceived affective costs (with 10,000 bootstrap samples). We created difference scores along the within-subjects factor to assess the indirect effects of

the between- by within-subjects interaction. The values in parentheses represent the total effects prior to the inclusion of the mediators. *Notes.* $^{\dagger}p < .15$, $*p < .05$, $**p < .01$, $***p < .001$.

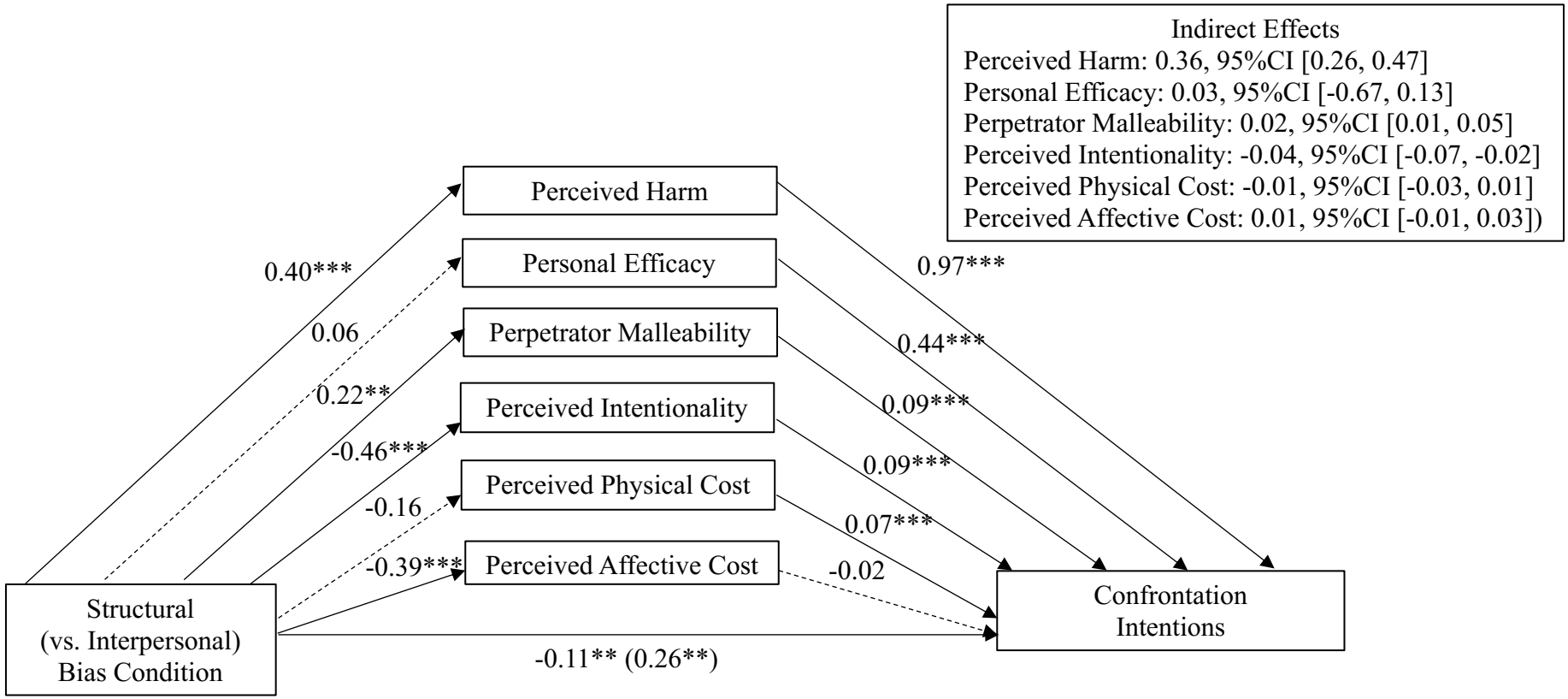

Figure 6: Study 3 process model of the indirect effect of bias condition (structural condition vs. interpersonal condition) on confrontation intentions via perceived harm, perceived perpetrator malleability, personal efficacy, perceived intentionality, perceived affective cost, and perceived physical cost (with 10,000 bootstrap samples). The values in parentheses represent the total effects prior to the inclusion of the mediators. *Notes.* $^{\dagger}p < .10$, $*p < .05$, $** p < .01$, $***p < .001$.