Does Confronting Prejudice Reduce Intergroup Bias? A Meta-Analytic Review

Chantelle Wood[1] https://orcid.org/0000-0001-6451-7859

Sofia Persson[2] https://orcid.org/0000-0002-7353-5204

Lilith Roberts[1] https://orcid.org/0000-0001-8654-2801

Oliver Allchin[3] https://orcid.org/0009-0000-4839-9725

Melanie Simmonds-Buckley[1] https://orcid.org/0000-0003-3808-4134

[1] School of Psychology, University of Sheffield, Sheffield, United Kingdom

[2] School of Social Sciences, Leeds Beckett University, Leeds, United Kingdom

[3] The University Library, University of Sheffield, Sheffield, United Kingdom

Correspondence concerning this article should be addressed to Chantelle Wood,

chantelle.wood@sheffield.ac.uk

Roberts: Conceptualization, Methodology, Investigation, Writing - Review & Editing; Oliver Allchin: Methodology, Investigation, Writing - Review & Editing; Melanie Simmonds-Buckley: Formal analysis, Data curation, Visualization, Writing - Review & Editing.

Abstract

Confronting prejudice is a promising strategy for reducing intergroup bias. The current meta-analysis estimated the effects of confronting prejudice on intergroup bias in the confronted person and examined the impact of potential moderators. Eligible studies measured intergroup bias in participants confronted vs. not confronted for intergroup bias. A three-level mixed-effects analysis on 91 effect sizes found a significant, medium-sized effect of confronting prejudice on reducing intergroup bias ($g_+ = 0.54$). There was only limited evidence of publication bias. Confrontation was differentially effective at reducing different types of intergroup bias with a medium-to-large effect on using or endorsing stereotypes, small-to-medium effects on behavior and behavioral intentions, and no significant effects on cognitive prejudice. Effects were otherwise largely robust to differences in confrontation, sample, and study design characteristics. Yet, studies predominantly focused on whether confronting the use of stereotypes reduced subsequent use of stereotypes in artificial settings, and primarily sampled US-based, young, White adults, making it difficult to generalize effects to other forms of intergroup bias and populations, particularly in real-world settings. Studies also tended to measure intergroup bias immediately after confrontation, so the duration of effects over longer periods is less clear. To better evaluate the potential of confrontation as a prejudice reduction technique, future research should examine whether confronting prejudice reduces different forms of intergroup bias in more diverse participant samples and settings, over longer periods, and further test theoretical mediators of these effects.

Public Significance Statement

This review shows that confronting prejudice reduces stereotyping and increases intentions to avoid expressing prejudice, though it does not change the prejudiced beliefs people report. Our results therefore show that confronting prejudice is a promising strategy for promoting fair and inclusive societies. Yet, studies primarily focused on young, White US-based adults and the immediate effects of confronting prejudice in laboratory or online settings. Research needs to evaluate whether confronting prejudice is effective in different groups of people in real-world settings over longer periods before applying these findings to practice.

Keywords: prejudice confrontation, bystander anti-prejudice, bystander anti-racism, prejudice reduction, intergroup bias, meta-analysis

Does Confronting Prejudice Reduce Intergroup Bias? A Meta-Analytic Review

A growing body of research investigates whether confronting prejudice – expressing "one's dissatisfaction with prejudicial and discriminatory treatment to the person who is responsible for the remark or behavior" (Shelton et al., 2006, p. 67) – can reduce prejudice and other forms of intergroup bias in the person confronted. For example, White people have been found to use fewer racial stereotypes (Chaney & Sanchez, 2018; Czopp et al., 2006), exhibit less racist attitudes (Czopp et al., 2006), and engage in less racist behavior (Munger, 2016) when confronted about racist transgressions by another White person. The effects of confrontation on reducing intergroup bias have been demonstrated for multiple targets of bias including racism (Burns & Monteith, 2019; Czopp et al., 2006; Lewis & Yoshimura, 2017) and sexism (Burns & Granz, 2021; Parker et al., 2018) and may continue to affect intergroup bias for days or weeks (Chaney & Sanchez, 2018; Chaney et al., 2021; Munger, 2016). Although there have been a number of narrative reviews investigating the effects of confronting prejudice, stereotyping, and discrimination on intergroup bias (e.g., Chaney et al., 2022; Monteith & Mark, 2005; Nelson et al., 2011), there has not yet been a comprehensive quantitative systematic review of this literature. Although Parker et al. (2018) report a meta-analysis of the effects of confronting prejudice on intergroup bias, their analysis focused on the effects and mediators of four studies within a single paper. Similarly, although Paluck et al.'s (2021) meta-analysis of prejudice reduction interventions included some studies evaluating the effects of confronting prejudice on reducing intergroup bias, they did not comprehensively cover all relevant research on confrontation, merged confrontation interventions with interventions relating to value consistency and self-worth for analysis, and also excluded studies that focused on gender-based prejudice. As a result, it is currently unclear whether confronting prejudice reduces intergroup

bias and when (i.e., under what conditions). The current paper therefore reports a comprehensive

and systematic review with meta-analysis focusing exclusively on prejudice confrontation,

seeking to estimate the overall effect size of confronting prejudice, stereotyping, or

discrimination on intergroup bias as well as investigate potential moderators of this effect.

Benefits of Prejudice Confrontation as a Strategy for Reducing Intergroup Bias

In Britain, 42% of people report having experienced prejudice in the preceding year, with

substantially higher rates from people in Minoritized groups, including Muslims (70%) and

people of a Black ethnic background (64%) (Abrams et al., 2018). In the United States, 44% of

people report experiencing discrimination either from time to time or regularly, with rates rising

to 63% amongst Minoritized groups (Lee et al., 2019). There is clearly substantial scope for

action to tackle intergroup bias.

A large number of effective strategies for reducing prejudice have been developed and

evaluated, including approaches based on intergroup contact, cooperative learning, peer

influence and discussion, diversity and anti-bias training, social categorization, cognitive and

emotional training, techniques for bolstering self-worth or value consistency, and entertainment-

based strategies (see Paluck et al., 2021, for a meta-analytic review). Although some strategies

for reducing prejudice occur naturally (e.g., spontaneous opportunities for intergroup contact or

cooperative learning), most are designed and delivered in a top-down manner, with the

associated resourcing challenges that brings. In contrast, as an individually driven action,

confrontation can operate without external guidance, and can be implemented as and when

intergroup bias is encountered, by both targets of prejudice and, perhaps more crucially, their

allies. This is especially relevant given the widespread attention of movements like Black Lives

Matter and #MeToo, where members of privileged groups are increasingly asking how they can best support Minoritized groups.

Confrontation can also help people cope with being targeted by prejudice, attenuating the negative effects of intergroup bias on psychological well-being (see Chaney et al., 2015, for a review). For example, Noh and Kaspar (2003) found that personal confrontation was one of the most common strategies used to cope with prejudice and discrimination by Korean adults living in Canada, and also the most effective strategy for reducing emotional distress. Similarly, Sanchez et al. (2016) found that more frequent confrontation of discrimination by racial minority students in the United States was associated with greater psychological well-being, and, according to Becker et al. (2014), women report increased self-esteem and confidence after confronting sexism. It is therefore clear that prejudice confrontation is an important strategy for reducing intergroup bias.

How Does Confronting Prejudice Reduce Intergroup Bias?

There are open questions about how confronting prejudice reduces intergroup bias. Monteith and colleagues' recent Validity and Impugnment as Determinants of Other-Confrontation Consequences theoretical framework (VIDOCC: Monteith et al., 2022) sets out potential mechanisms that may underlie the effects of confronting prejudice on intergroup bias. Incorporating the model of self-regulation of prejudice (e.g., Monteith, 1993; Monteith et al., 2002), the VIDOCC theoretical framework argues that confronting prejudice can promote awareness of discrepancies between personal standards and actual expressions of prejudice (i.e., prejudice-related discrepancies) in the person confronted, resulting in negative self-directed affect, and reflection or rumination, followed by inhibition of the biased response. The personal standards on which prejudice-related discrepancies are predicated can be based not only on one's

own viewpoint but also on the viewpoints of others (i.e., social norms). As such, the VIDOCC

framework incorporates previous theorizing that confrontation may reduce expressions of

intergroup bias by communicating or strengthening social norms about the (un)acceptability of

prejudice (i.e., injunctive social norms) and/or its (in)frequency (i.e., descriptive social norms),

increasing the likelihood that the confronted person will modify their behavior to conform to

anti-prejudiced norms. For example, research indicates that people report lower prejudice

(Crandall et al., 2002) and are less likely to endorse negative stereotypes (Stangor et al., 2001)

when negative feelings or beliefs about a target group are seen as less socially acceptable.

Similarly, evidence suggests that people express stronger anti-racist attitudes when exposed to a

confederate who expresses vs. does not express anti-racist attitudes (Blanchard et al., 1994).

Monteith et al. (2022) also stipulate a critical boundary condition for effects on prejudice

reduction in the VIDOCC framework: the validity of the confrontation. Specifically, confronting

prejudice should only prompt people to consider potential prejudice-related discrepancies if the

confrontation is perceived as valid (i.e., if people agree that they have indeed behaved in a biased

way). Confrontations that are not perceived as valid should not result in a reduction in intergroup

bias.

As the VIDOCC (Monteith et al., 2022) is a newly published theoretical framework, there

is currently no research that directly measures whether perceiving a confrontation as valid is

necessary for confrontation to reduce intergroup bias (though see Burns & Granz, 2021; Gulker

et al., 2013 for measures of acceptance of confrontation). Conversely, in line with the

mechanisms implicated in the VIDOCC, research by Monteith and colleagues (see Monteith &

Mark, 2005, for a summary) has demonstrated that confrontation is associated with greater

negative self-directed affect, such as feelings of guilt, regret, shame and anger (Chaney &

Sanchez, 2018; Czopp et al., 2006; Parker et al., 2018), and greater rumination (Burns & Granz, 2021; Chaney & Sanchez, 2018). There is also building evidence that negative self-directed affect and rumination mediate the effect of confronting prejudice on reducing intergroup bias (Burns & Granz, 2021; Chaney & Sanchez, 2018; Czopp et al., 2006; Parker et al., 2018).

Conversely, evidence that confrontation strengthens social norms is more limited and less consistent. Recent research by Li et al. (2024, Study 2) found that non-Hispanic/Latinx participants perceived stronger injunctive and descriptive anti-bias social norms when they observed a White person confronting bias towards Mexican job candidates, relative to when bias was not confronted. In contrast, when Li et al. examined anti-Black bias in a similar scenario with both Black and White participants (Study 3), only Black participants showed the same effects. In addition, confrontations that emphasize social norms against the expression of prejudice (e.g., Burns & Monteith, 2019; Parker et al., 2018) appear no more effective in reducing intergroup bias than confrontations that are less explicitly norm-based. Similarly, individual differences in external motivation to control prejudice (i.e., to conform to social norms regarding the expression of prejudice) have not been found to moderate the effects of confrontation on intergroup bias (Burns & Monteith, 2019).

The relatively small size of the above literature means that at the current time, it is not possible to conduct a direct meta-analytic test of the mediators proposed to underlie the effects of confronting prejudice in the VIDOCC (Monteith et al., 2022) or evaluate the validity of the confrontation as a key boundary condition. Instead, the current meta-analysis aims to guide future development and evaluation of theory, by highlighting where the moderators we test are relevant to the key mechanisms implicated in the VIDOCC.

Moderators of the Effect of Confrontation of Prejudice on Intergroup Bias

We aimed to examine 21 moderators that could shed light on factors that influence the effect of confronting prejudice on intergroup bias; this was intended to guide future empirical work and interventions that seek to increase the effectiveness of confrontation as a prejudice reduction technique. The following 19 pre-registered moderators were extracted and/or coded: characteristics of the confrontation (the group membership of the confronter relative to participants, the identity of the confronter, the target of confrontation, the target of intergroup bias, whether the confrontation was performed in public or private), characteristics of the measure of intergroup bias (type of intergroup bias measured, implicit vs. explicit measures of intergroup bias), characteristics of the sample (mean age, gender composition, ethnicity composition, sexual orientation composition, sample type, country of participant recruitment), characteristics of the study design (confrontation modality, study type, time between confrontation and measurement of intergroup bias), and characteristics of the publication (publication year, publication status, publication language). We also examined the effect of confrontation style and the author's laboratory, which were not pre-registered and therefore considered exploratory. The moderator 'confrontation modality' was originally called 'type of confrontation' in our pre-registration, but was changed to better distinguish this moderator from confrontation style.

Characteristics of the Confrontation

Group Membership of the Confronter Relative to Participants. According to the VIDOCC theoretical framework (Monteith et al., 2022), one route by which confronting prejudice can reduce intergroup bias is through raising awareness of discrepancies between social norms regarding the expression of bias and actual prejudiced responses. Although the

VIDOCC framework makes no assumptions about which sources of social norms are more influential, self-categorization approaches to social influence (e.g., Abrams & Hogg, 1990) argue that people should be more likely to conform to ingroup vs. outgroup norms. In line with this argument, Stangor et al. (2001) found that people reported more positive attitudes towards African Americans when they were told that ingroup members endorsed more positive stereotypes, relative to when they were told that outgroup members endorsed more positive stereotypes. In relation to confrontation, however, research is equivocal. On the one hand, Meyers et al. (2020) found that Asian participants rated racist social media posts as more offensive when the posts were first confronted by a member of their ingroup (Asian person) vs. outgroup (White person). On the other hand, Li et al. (2024) found that although Black participants perceived stronger anti-racist social norms when they observed confrontation of racial bias by an outgroup member (White person), relative to no confrontation, White participants observing the same scenarios showed no effect of confrontation on social norms, despite the confronter being an ingroup member. It is therefore unclear whether we would expect confrontations to be more effective at reducing intergroup bias when they are delivered by someone who shares the relevant ingroup identity with the person confronted (e.g., gender, if sexism is being confronted; race, if racism is being confronted) relative to when delivered by someone who does not share that relevant ingroup identity. Accordingly, in the current meta-analysis, we examine whether the group membership of the confronter relative to participants (ingroup; outgroup; irrelevant, mixed, or unspecified) moderates the magnitude of the effect of confronting prejudice on intergroup bias.

Identity of the Confronter. A core advantage of confrontation as a prejudice reduction strategy is its capacity to be employed by both members of the group targeted by prejudice and

by non-members of that group (i.e., allies). Yet, whether targets and allies differentially reduce intergroup bias through confrontation is the subject of continued empirical debate. On the one hand, research suggests that confrontations performed by members of the group targeted by prejudice may be taken less seriously: Confrontations by members of the target group are less accepted than confrontations by allies (Gulker et al., 2013), and confronters from the target group are more likely to be seen as overreacting (Czopp & Monteith, 2003; Kutlaca et al., 2020) and more likely to be derogated or characterized as complainers (Eliezer & Major, 2012; Gulker et al., 2013) compared to allies. Monteith et al. (2022, p. 36) suggest that these effects may be driven by the perception that members of the target group have a "vested interest in bias reduction", reducing their apparent objectivity and thus the perceived validity of the confrontation. If the effects of confrontation on intergroup bias are contingent on the confrontation being perceived as valid (i.e., if the tenets of the VIDOCC hold true), we might therefore expect target group confronters to be less effective than ally confronters.

On the other hand, these differences in how confrontations are perceived when performed by targets vs. allies are not consistently followed by differential effects on intergroup bias. Although Munger (2016) found that Twitter users only reduced their subsequent use of racial slurs when confronted for racist tweets by an ally (vs. a member of the target group), other studies have found no evidence that confronter identity moderates post-confrontation reductions in stereotyping (Chaney et al., 2021; Czopp et al., 2006). Wallace et al. (2024, p. 3) argue that such mixed results may occur because members of the group targeted by prejudice are seen as "biased but expert". In their research, participants rated female confronters of gender bias as both more biased and more expert relative to male confronters. As such, although a target's 'vested interest' may attenuate the effect of confronting prejudice on intergroup bias, the expertise

conferred by lived experience may increase the effect, resulting in an overall null effect of confronter identity.

To further complicate the discussion, some studies instead manipulate confrontation through computerized feedback highlighting the use of stereotypes or biased decision-making during computer-based tasks (e.g., Burns & Monteith, 2019; Parker et al., 2018). Although computerized feedback on intergroup bias is arguably objective, research indicates that people often respond defensively to this feedback, particularly when it reveals biases that diverge from consciously held attitudes (see Howell et al., 2024 for a review). No research to date examines how the effects of computer-mediated forms of confrontation of prejudice might compare to interpersonal forms.

The current meta-analysis therefore tests whether the identity of the confronter (actual target of intergroup bias, same group as target of intergroup bias, different group from target of intergroup bias [i.e., ally], computer [i.e., non-interpersonal confrontation], or mixed/unspecified group) moderates the magnitude of the effect of confronting prejudice on reducing intergroup bias.

Target of Confrontation. People can either directly experience prejudice confrontation (e.g., when they are confronted) or observe prejudice confrontation (e.g., when another person is confronted). Although the VIDOCC theoretical framework (Monteith et al., 2022) focuses on explaining directly experienced confrontation, it would be both theoretically interesting and practically useful to examine whether effects extend to observed confrontation. That is, if observing confrontation of prejudice shows similar effects to experiencing confrontation of prejudice, the VIDOCC may usefully be expanded to encompass and explain such effects. Such findings would also suggest relative practical benefits of 'calling out' prejudice in public, as

effects on reducing intergroup bias would then not be limited to the person confronted, but also extend to the bystanders present. Although the evidence is very limited, research suggests that the outcomes proposed to promote control of prejudiced responses in the VIDOCC (i.e., negative affect) are attenuated if a person observes rather than directly experiences prejudice confrontation. Hyers (1999), for example, found higher levels of guilt in people who were directly confronted for overtly prejudicial comments relative to people who did not outwardly express prejudice and who therefore only observed confrontation. Although existing research has looked at the effects of both witnessed confrontation (e.g., Boysen, 2013) and experienced confrontation (e.g., Chaney et al., 2021; Czopp et al., 2006; Lewis & Yoshimura, 2017; Parker et al., 2018; Simon et al., 2012) on intergroup bias, to our knowledge no research has directly manipulated and compared the two in a single study. The current meta-analysis therefore aims to test whether the target of confrontation (experienced confrontation, observed confrontation, mixed confrontation) moderates the magnitude of the effect of confronting prejudice on reducing intergroup bias.

Target of Intergroup Bias. Research on confrontation of prejudice has focused on a limited range of target groups, predominantly examining the effects of confrontation on intergroup bias towards Black people (e.g., Burns & Monteith, 2019; Chaney & Sanchez, 2018; Chaney et al., 2021; Czopp et al., 2006; Munger, 2016; Parker et al., 2018) or women (e.g., Boysen, 2013; Burns & Granz, 2021; Burns & Monteith, 2019; Parker et al., 2018). In contrast, few studies look at the effects of confrontation on intergroup bias towards other ethnicities (e.g., Lewis & Yoshimura, 2017) or Minoritized identities (e.g., gay people: Hyers, 2010). Although the VIDOCC theoretical framework (Monteith et al., 2022) does not incorporate the target of bias as a moderator of the effects of confrontation on intergroup bias, its authors argue that

confrontation of racism is more likely to be considered valid than confrontation of sexism. For example, sexism is less likely to be seen as worthy of confrontation relative to racism (Woodzicka et al., 2015). There is also evidence relevant to other mechanisms implicated in the VIDOCC framework: Some studies have found that being confronted for sexism has less of an impact on self-directed negative affect such as guilt, discomfort and depressed affect (Czopp & Monteith, 2003), and there is evidence of weaker social norms denouncing sexism relative to racism. Specifically, sexist statements, jokes, or hate speech are perceived as less offensive than their racist equivalents (Cowan & Hodge, 1996; Woodzicka et al., 2015), confrontation of sexism is seen as more amusing than confrontation of racism towards Black people (Czopp & Monteith, 2003), and on average, external motivation to respond without prejudice is weaker for sexism vs. racism towards Black people (Klonis et al., 2005).

Despite the above findings, the limited number of studies that compare the confrontation of gender vs. racial bias have found no evidence that subsequent effects on intergroup bias are moderated by the target of intergroup bias (e.g., Burns & Monteith, 2019; Parker et al., 2018). The current meta-analysis therefore tests whether the target of intergroup bias (Black people, women, sexual minorities, other ethnic identities, religious identities, weight status, mixed groups [different target groups within the same comparison]) moderates the magnitude of the effect of confronting prejudice on reducing intergroup bias.

Confrontation Style and Whether Confrontation is Performed in Public or Private. A variety of approaches have been used to categorize and explore the impact of different styles of confronting prejudice on reductions of intergroup bias. Research has tended to compare responses to confrontations that are more or less threatening or hostile (e.g., Becker & Barreto, 2014; Czopp et al., 2006; Hyers, 2010), though some research suggests that at least in

hypothetical scenarios, allies are more likely to respond with educational or value-based arguments centering around morality and empathy, than more aggressive confrontations (Hubbard et al., 2013). In general, although people feel more positively about less threatening or less hostile confrontations and confronters (Becker & Barreto, 2014; Martinez et al., 2017), there is no clear evidence that these types of confrontations are differentially effective at reducing subsequent stereotyping (Czopp et al., 2006; Hyers, 2010).

Chaney and Sanchez (2021) identify five styles of prejudice confrontation: Educational, argumentative, help-seeking, empathy, and humorous. Research conducted to validate the associated Prejudice Confrontation Styles (PCS) scale suggests that educational, argumentative, and empathetic confrontation styles are perceived as more effective at changing the future behavior of the perpetrator than humorous and help-seeking styles (Chaney & Sanchez, 2021), though other research suggests that humorous and non-humorous confrontations are perceived as equally effective (Woodzicka et al., 2020). A good understanding of which types of confrontation are most effective is critical for guiding future research and practice. The current meta-analysis therefore examines whether the effect of confronting prejudice on reducing intergroup bias is moderated by confrontation style, as operationalized in the PCS scale (argumentative, educational, empathy, humorous, help-seeking, other). As this scale was published shortly after we pre-registered our meta-analysis, these analyses were not pre-registered and are thus exploratory.

It may also be important to consider whether the confrontation occurs in a public or private setting. Activists' and laypersons' discourse often distinguishes between confrontations that 'call-out' vs. 'call-in' intergroup bias (Mahan, 2017; Woods & Ruscher, 2021). Call-outs shame perpetrators for their bias in public whereas call-ins are made in private and couched in

compassion (Mahan, 2017). Despite anecdotal criticism of contemporary 'call-out culture',

recent results from a survey in the United States indicate relatively prevalent positive attitudes to

public confrontation (Vogels et al., 2021). But are public and private confrontations equally

effective at reducing prejudice? To our knowledge, no studies have yet examined whether public

and private confrontations are differentially effective at reducing intergroup bias. The VIDOCC

theoretical framework (Monteith et al., 2022) makes no explicit predictions about the differential

effects of public vs. private confrontations, and our own conjecture based on our understanding

of this framework is equally ambivalent. On the one hand, we might speculate stronger effects of

public relative to private confrontation, because public confrontations may provide sterner

evidence for anti-prejudiced social norms (i.e., because the confronter is willing to confront in

the presence of other people) than private confrontations. On the other hand, as pointed out by an

anonymous reviewer, people may also become more defensive in public contexts and more

motivated to deny the validity of the confrontation. The current meta-analysis therefore tests for

the first time whether the effect of confronting prejudice on reducing intergroup bias is

moderated by whether the confrontation was performed in public vs. private.

Characteristics of the Measure of Intergroup Bias

Type of Intergroup Bias Measured. Most research on the confrontation of prejudice

examines the effect of confronting the use of stereotypes on subsequent use of stereotypes. Use

of stereotypes is typically measured using an inference task, where participants are given

photographs of target group members accompanied by short descriptions, and asked to make

inferences about the people pictured (e.g., Burns & Granz, 2021; Burns & Monteith, 2019;

Chaney & Sanchez, 2018; Czopp et al., 2006; Lewis & Yoshimura, 2017). On key trials, the

descriptions can be interpreted in either stereotypic or non-stereotypic ways (e.g., "This person

can be found behind bars" can be interpreted as stereotypic [criminal] or non-stereotypic

[bartender] of Black people: Czopp et al., 2006, p. 788), and thus participants' responses on these

trials can be coded as either stereotypic or non-stereotypic. The task is completed twice—at the

first time point, the tendency for people to respond in stereotypic ways to key trials is used as the

basis for confrontation ("...i *[sic]* noticed that for some of the pictures of Black people you said

some stereotypical things like criminal, bum, and welfare": Czopp et al., 2006, p. 792), and at the

second time point, after confrontation, the number of stereotypic responses forms the measure of

stereotype application. Although some studies have looked at the impact of confrontation on

other measures of intergroup bias, such as attitudes (Boysen, 2013; Czopp et al., 2006; Hyers,

2010), behavioral intentions (Burns & Granz, 2021; Parker et al., 2018), and behavior (Chaney

& Sanchez, 2018; Munger, 2016), most still manipulate confrontation based on the use of

stereotypes in the inference tasks described above (e.g., Burns & Granz, 2021; Chaney &

Sanchez, 2018; Czopp et al., 2006) or other scenarios (Boysen, 2013). Given insufficient

variation in the type of intergroup bias confronted, it is not possible for the current meta-analysis

to examine whether confrontation of different types of intergroup bias has differential effects.

Instead, we explore whether the type of intergroup bias *measured after* confrontation (affective

prejudice, cognitive prejudice, stereotyping, behavioral intentions, behavior, other) moderates the

effect of confronting prejudice on reducing intergroup bias.

Implicit vs. Explicit Measures of Intergroup Bias. Research on the confrontation of

prejudice overwhelmingly measures explicit (e.g., stereotype use: Burns & Granz, 2021; Chaney

& Sanchez, 2018; Czopp et al., 2006) rather than implicit forms of bias. This reflects theoretical

explanations for the potential impact of confronting prejudice on intergroup bias, such that

confrontation is assumed to promote conscious self-regulation of future behavior to better adhere

to personal standards. In the current meta-analysis, we therefore sought to examine whether the effect of confronting prejudice on reducing intergroup bias was moderated by whether the measure of intergroup bias was explicit vs. implicit.

Characteristics of the Study Design

Confrontation Modality and Study Type. As confrontation is challenging to manipulate and control in experimental settings, the majority of research examining the impact of confrontation of prejudice on intergroup bias has implemented confrontation in written modalities; for example, through computer-mediated typewritten conversations (e.g., Czopp et al., 2006; Lewis & Yoshimura, 2017), or computerized feedback on a task (e.g., Burns & Granz, 2021; Parker et al., 2018). A smaller number of studies have used confederates or actors (e.g., Hyers, 2010) or the experimenter (e.g., Chaney & Sanchez, 2018; Chaney et al., 2021) to engage in verbal confrontations. For similarly pragmatic reasons, most confrontation research involves lab-based or online studies (e.g., Boysen, 2013; Burns & Granz, 2021; Czopp et al., 2006; Lewis & Yoshimura, 2017; Parker et al., 2018) with very few studies using field-based approaches (e.g., confrontation via twitter: Munger, 2016). No research has examined whether confrontation modality (verbal, written) or the type of study (laboratory, field, online) moderates the effect of confronting prejudice on reducing intergroup bias. Therefore, to guide future empirical work, we test both potential moderators in the current meta-analysis.

Time Between Confrontation and Measurement of Intergroup Bias. Research on confronting prejudice has only recently turned to examining the persistence of effects over time, from up to a week later (e.g., Burns & Monteith, 2019; Chaney & Sanchez, 2018; Chaney et al., 2021) to multiple weeks (e.g., Munger, 2016). As a result, there is a lack of research that compares the effects of confrontation at different time points. The few existing studies suggest

that confronting prejudice significantly reduces intergroup bias up to a few days (Burns &

Monteith, 2019) and weeks later (Munger, 2016), but is not effective after 2 months (Munger,

2016). In the current meta-analysis, we therefore examine the potential decay of effects of

confronting prejudice on reducing intergroup bias, by examining whether the time between

confronting prejudice and measurement of intergroup bias (in days) moderates the magnitude of

the effect.

Characteristics of the Sample

Prejudice reduction interventions are predominantly evaluated in student populations

(Paluck et al., 2021), despite evidence that these interventions may be less effective in other adult

populations (e.g., see Hsieh et al., 2022). To place the results of the current meta-analysis in the

context of the samples recruited by the primary studies and to quantify constraints on generality

in line with current recommendations (see Simons et al., 2017), we report summary statistics for

the following commonly reported demographic characteristics of the included participant

samples: mean age, gender composition, ethnicity composition, sexual orientation composition,

sample type, and country of participant recruitment, and also note where these data are not

reported. Where there are sufficient available data or variation in data, we also examine whether

these characteristics moderate the effect of confronting prejudice on reducing intergroup bias.

Characteristics of the Publication

Finally, we sought to examine four characteristics of the publication: publication year,

publication status, author's laboratory, and publication language. First, we explore whether there

is a 'decline effect' (e.g., see Schooler, 2011) in the prejudice confrontation literature, by

examining whether publication year moderates the effect of confronting prejudice on reducing

intergroup bias. Second, we examine whether publication status moderates the effect of

confronting prejudice on reducing intergroup bias, to provide an additional test of potential

publication bias (i.e., larger effect sizes reported in published vs. unpublished work) alongside

the other tests of publication bias described in the method section. Third, we compared effect

sizes on papers where authors are (currently or historically) linked with Monteith's laboratory,

Chaney's laboratory, or other laboratories, to explore whether key laboratories (classed as those

producing 2 or more papers) have developed particularly effective methods for confronting

prejudice or measuring its impact. We did not expect any significant moderating effects as a

function of author's laboratory. This moderator was suggested by an anonymous reviewer and

was not pre-registered. The analysis should therefore be considered exploratory. Finally, we

planned to examine whether effect sizes were moderated by publication language.

The Present Research

Despite the growing body of research evaluating the effects of confronting prejudice on

intergroup bias, there is currently no clear consensus regarding whether and how different

characteristics of the confrontation or intergroup bias (including how bias is measured) moderate

the size and direction of the effect, and whether effects are relatively robust to differences in

study design or sample characteristics. The present meta-analysis aimed to address these gaps to

support theory development and provide guidance on how to optimize the effectiveness of

confrontation of intergroup bias as a prejudice reduction strategy. Specifically, we estimate the

overall effect of confronting prejudice on intergroup bias and evaluate factors that may influence

the size and direction of the effect.

<div align="center">

Method

</div>

Transparency and Openness

PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses: Page et

al., 2021), MARS (Meta-Analysis Reporting Standards: Appelbaum et al., 2018), and TOP

(Transparency and Openness Promotion) guidelines were followed when conducting and reporting the meta-analysis. The meta-analysis protocol (including search terms and strategies, inclusion criteria, data screening, data extraction and moderator coding, risk of bias assessment, and data synthesis strategies) was initially pre-registered on OSF in April 2021. The pre-registration, data and analysis syntax are available on the Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/2X3BY). One additional moderator (group membership of the confronter relative to the participant) was pre-registered in May 2023, before beginning coding of this moderator. Two other moderators (confrontation style, author's laboratory) were not pre-registered and thus should be considered exploratory. We did not pre-register any predictions regarding the results of moderator analyses. Search terms were updated in March 2023 with support from a university librarian (the fourth author). To avoid duplication of effort, records from the amended searches were cross-referenced with screening results and data extraction from the original searches, such that screening results and data extraction from records common to both searches were pulled across to the amended search. In addition, we cross-referenced the original and amended searches to ensure that all records screened as meeting the inclusion criteria at the title and abstract screening stage in the original searches, were either present or manually added to the amended search results. Finally, in order to address the statistical interdependency of effect sizes in our meta-analysis, we chose to employ a three-level meta-analytic strategy (Viechtbauer, 2010) to analyze our data rather than the pre-registered two-level random-effects meta-analysis. Further details can be found in the section on Effect Size Multiplicity and the section on Analysis Strategy. Any other deviations from the pre-registered protocol are identified as relevant below.

Literature Searches

Searches for published and unpublished literature were conducted using Scopus, Web of Science, APA PsycInfo, and OSFPreprints by the fourth author. The Web of Science search covered the following databases: Web of Science Core Collection, BIOSIS Citation Index, BIOSIS Previews, Current Contents Connect, Data Citation Index, Essential Science Indicators, Journal Citation Reports, KCI-Korean Journal Database, MEDLINE, SciELO Citation Index, and Zoological Record. Searches were originally run on the 13th of March 2023 and updated on the 3rd of January 2024. The search strategies for each database and platform can be found on OSF and were structured so that a record had to contain at least one term relating to intergroup bias and at least one term relating to confrontation. Search terms for intergroup bias were as follows: prejudic*, anti-prejudic*, antiprejudic*, "prejudice reduction", racis*, anti-racis*, antiracis*, sexis*, anti-sexis*, antisexis*, homophob*, anti-homophob*, antihomophob*, islamophob*, anti-islamophob*, anti-muslim, antimuslim, ableis*, anti-ableis*, antiableis*, ageis*, anti-ageis*, antiageis*, discriminat*, stereotyp*, stigma*, "intergroup bias", anti-fat, and antifat. Search terms for confrontation were as follows: confrontation, "bystander intervention", "bystander action", "ally intervention", "ally action", and "bystander anti*". These searches resulted in 3584 records, 1204 of which were removed as duplicates, leaving 2380 records.

Eleven records from the superseded pre-registered searches (last updated on the 21st of April 2022) that had been screened as meeting the inclusion criteria at the title and abstract screening stage were added to the records from the amended searches.

Requests for studies were placed on the Society for Personality and Social Psychology Open Forum and the European Association of Social Psychology Mailserv on the 14th of May 2021, and Twitter (now X) on the 14th of May 2021, and 15th of March 2023, resulting in contact

from 5 authors. Lead authors of papers on the confrontation of prejudice that were identified in

the above searches for published literature and other researchers suggested by those authors ($N =$

63) were also contacted to request any unpublished studies or studies in preparation. Authors

were initially contacted by the first author between the 11th of May 2021 to the 24th of November

2021 after original searches were run, and where relevant, contacted again or for the first time

between the 13th of March 2023 to the 28th of March 2023 after amended searches were run. 44

of 56 authors (79%) replied to the request in the first round of emails, and 30 of 61 (49%) replied

to the second round of emails. Overall, contact with authors resulted in 22 new records and 1

record flagging up supplemental studies linked to an existing record. Figure 1 shows the

PRISMA 2020 (Page et al., 2021) literature search flow diagram.

Inclusion and Exclusion Criteria

In order to be included in the meta-analysis, studies needed to have an experimental

design with 1) an intervention group who were either a) confronted for expressing or

demonstrating intergroup bias or b) observed another person or people being confronted for

expressing or demonstrating intergroup bias; 2) a control group who were not confronted for

expressing or demonstrating intergroup bias, or who did not observe others being confronted;

and 3) a post-intervention (i.e., post confrontation or alternative control task) measure of

intergroup bias in both the intervention and control group. Measures of intergroup bias included

(but were not limited to) those measuring cognitive (e.g., stereotyping), affective (e.g.,

prejudice), and behavioral (e.g., discrimination) components of bias, along with overarching

constructs (e.g., stigma). There were no restrictions on publication status or date. There were no

inherent restrictions on language, except for the stipulation that only records that could be

adequately translated into English (e.g., through digital applications or services at the host institution) would be included in the meta-analysis.

Although not made explicit in the pre-registered protocol, manipulations involving imagined confrontation (e.g., Hillard, 2011) or self-confrontation (e.g., Kennedy, 1995) did not meet the inclusion criteria of being confronted or observing a confrontation for expressing or demonstrating intergroup bias. Self-generated confrontations seem unlikely to be driven by the same theoretical processes (e.g., raising awareness of prejudice-related discrepancies) as confrontations by others, and the number of studies evaluating these strategies was too low to permit moderator analysis (i.e., fewer than five effect sizes). Intervention groups that were described as involving confrontation of intergroup bias, but that did not unambiguously do so (i.e., "... I wonder if some people would view some of your answers as warmed over!": Off-record confrontation, Lewis & Yoshimura, 2017) were excluded. Three records included intervention or control groups (Burns & Monteith, 2019, Study 3; Parker et al., 2018, Study 4; Simon & O'Brien, 2015) where participants were told (or it was implied) that they had shown no intergroup bias. Given the potential for this manipulation to contaminate the effects of confrontation, these conditions were excluded from the meta-analysis. One record (Mark, 2007) manipulated social norms alongside confrontation such that there were three intervention groups (confrontation + sexist social norms, confrontation + non-sexist social norms, confrontation only) and two control groups (no confrontation + sexist social norms, no confrontation + non-sexist social norms). Although the confrontation-only group represents the 'purest' and thus most ideal manipulation of confrontation, there was no equivalent control group (i.e., all control groups manipulated social norms), and accordingly, it would not be possible to disentangle effects driven by confrontation from those driven by the social norms manipulation in the control

groups. As a result, we excluded the confrontation-only condition, and reported data from the

confrontation and no-confrontation groups, collapsing over the social norms condition. Another

two records manipulated trust or affiliative motivation towards the confronter, orthogonally to

confrontation (Hildebrand et al., 2024, Study 2; Patrianakos, 2021). Given the potential for these

manipulations to contaminate the effects of confrontation, for Hildebrand et al. (2024), we

extracted data from the trust-neutral confrontation and trust-neutral no-confrontation control

groups only. Patrianakos (2021) manipulated affiliative motivation as high or low, and as such,

there was no clear affiliative motivation 'neutral' condition. Accordingly, we extracted data from

the racial confrontation and rude confrontation (control) conditions, collapsed over affiliative

motivation condition.

Measures of intergroup bias reflecting societal rather than personal biases (e.g., semantic

differentials reflecting perceptions of the cultural stereotype of Native Americans: Lewis &

Yoshimura, 2017) or bias towards groups that were not the target of the confrontation

intervention (e.g., studies that looked at transfer effects of confrontation of prejudice towards one

group on prejudice towards another group: Chaney et al., 2021, Study 3 and Supplemental Study

2) were not considered to meet the inclusion criteria. Studies measuring conative components of

intergroup bias (i.e., behavioral intentions) were included. Two outcome variables (the emotional

tone of a 'day in the life of' narrative written by participants about a target group member:

Kroeper, 2020; the detection of sexist language use in a series of sentences: Hildebrand et al.,

2024; Mallett & Wagner, 2011) were initially screened by both raters as meeting the inclusion

criteria for measures of intergroup bias. After further discussion, we decided that neither

outcome variable adequately measured intergroup bias, and thus they were excluded from the

analysis. Our reasoning was that the coding for the emotional tone of narratives written by

participants about a target group member did not distinguish between negative tone reflecting

intergroup bias (e.g., use of negative stereotypes) and negative tone reflecting acknowledgment

of intergroup bias (e.g., a narrative acknowledging experiences of prejudice and discrimination

in the daily life of the target group member). Accordingly, the data are difficult to interpret in

terms of intergroup bias. Similarly, although the use of sexist language would be considered a

measure of intergroup bias, the sexist language detection task instead measures participants'

ability to detect sexist language when *explicitly instructed to do so*. As a result, we do not

consider this task to unequivocally reflect intergroup bias.

Finally, where multiple reports of the same studies existed in different formats (i.e., a

PhD thesis and subsequent published paper), individual studies were included only once to avoid

duplication. Precedence was given to published reports of studies, except where the unpublished

records contained a more comprehensive record of data needed for the calculation of relevant

effect sizes or coding of moderators.

Effect Size Multiplicity

Strategies for dealing with multiple effect sizes from overlapping participant samples,

known as 'multiplicity' (Higgins et al., 2021; López-López et al., 2018) were overlooked in the

pre-registered protocol. During data extraction, it became clear that most studies reported

multiple eligible effect sizes for overlapping participant samples (i.e., reported multiple measures

of intergroup bias or multiple timepoints for the same participant sample, or compared multiple

experimental groups to a single control group). Unless their interdependency is accounted for,

the inclusion of multiple effect sizes from overlapping participant samples would violate

assumptions of independence of observations in the meta-analysis, which would lead to an

underestimation of the error variance associated with each effect size (Borenstein, 2009). We

therefore dealt with effect size multiplicity in two ways. First, in the interests of parsimony, where data from two control groups were reported (i.e., Czopp et al., 2006, Study 3; Lewis & Yoshimura, 2017), only the 'no confrontation' control group (the most commonly used control in the confrontation literature) was included in the analysis. This decision was made on the basis that a) inclusion of multiple control groups is not relevant to moderator analysis, and b) variation in the nature of the control group is unlikely to change the outcome of the meta-analysis. Second, retention of multiple outcome measures and time points was considered critical for moderator analyses, and as such, strategies that involved selecting one effect to include or collapsing dependent effect sizes into a single average effect size (referred to as reductionist strategies: López-López et al., 2018) were not considered appropriate. Accordingly, rather than the traditional two-level random-effects meta-analysis outlined in the pre-registered protocol, we elected to use a three-level meta-analytic strategy (Viechtbauer, 2010) that would allow us to include all relevant effect sizes while accounting for the statistical interdependency of effect sizes within individual studies. Further details are provided in the analysis strategy section.

Screening Process

In the first stage of screening, titles and abstracts of 2413 records were screened according to the following inclusion criteria: that the record focused on the confrontation of intergroup bias, reported empirical research, reported at least one study that employed eligible intervention and control groups, and included at least one eligible post-intervention measure of bias in the intervention and control groups. Screening was terminated for each study at the first point where an inclusion criterion was not met. All records obtained through literature searches were screened by the third author and all records obtained from alternative sources (e.g., contact with authors) were screened by the first author. Ten percent of the combined pool of records

(243 records) were also independently screened by the alternate author (i.e., the first author for records obtained through literature searches and the third author for records obtained from alternative sources). Interrater reliability for the outcomes of the first stage of screening (i.e., include vs. exclude) indicated almost perfect agreement ($\kappa = .86$), and discrepancies were resolved through discussion.

In the second stage of screening, 88 records were sought for retrieval. After the removal of one record and an additional study from another record that reported duplicate data, as well as one record that could not be accessed, the remaining 86 records comprising 169 studies were screened for eligibility based on the full text. Each study in each record was screened according to the same criteria as in the first stage of screening. One hundred percent of these records were screened independently by both the first and third authors. Interrater reliability for the Stage 2 screening outcome (i.e., include vs. exclude) indicated almost perfect agreement ($\kappa = .84$), and discrepancies were resolved through discussion. Twenty-three records comprising 38 studies were initially screened as meeting the inclusion criteria. One further study and one full record comprising one study were excluded after further discussion during data extraction (see section on inclusion and exclusion criteria for further detail), leaving 22 records comprising 36 studies suitable for inclusion. Records included in the meta-analysis are marked with an asterisk in the References section.

Data Extraction

Statistics for calculating effect sizes were independently extracted by the first and third authors. Specifically, the mean, standard deviation, and sample size for each post-confrontation measure of intergroup bias were extracted for both the confrontation and control conditions, along with information on whether a higher score on each measure of intergroup bias indicated

higher or lower bias. If means and standard deviations were not reported, the univariate *F*-value

and sample size for the confrontation and control conditions were instead used to calculate effect

size. Interrater reliability for each statistic was excellent (ICCs ranged from .997 – 1) and

discrepancies were resolved through discussion. The first author additionally extracted

information on whether multiple effect sizes within the same study came from overlapping vs.

independent samples of participants. In contrast to our pre-registration, we extracted means and

standard deviations for a difference score representing the change in bias from pre- to post-

confrontation for two records (Kroeper, 2020, Studies 2 and 3; Lewis & Yoshimura, 2017),

rather than extracting data solely post-confrontation. This was because Lewis and Yoshimura

(2017) were unable to provide the means and standard deviations for their post-confrontation

measure of stereotypic responses, so we chose to extract the data for the stereotypic response

change score rather than exclude this study. Kroeper's (2020) measure of stereotype usage asked

participants to write a story about a 'day in the life' of a target group member prior to the

confrontation manipulation, and then revise that same story after confrontation – the difference

in the number of stereotypes used in each story was then calculated. Given that post-

confrontation usage of stereotypes on the task is inherently constrained by pre-confrontation

usage of stereotypes, we decided that extracting data for the change score was most appropriate.

Where studies were eligible for inclusion but did not contain sufficient information for

effect sizes to be calculated or coding of moderators (35 studies from 21 records, by 15 authors),

authors were contacted by the first author to request this additional information between May

2021 to March 2023. 12 of the 15 authors (80%) responded and 9 of the 15 (60%) were able to

provide the requested data. Where calculation of effect sizes was prevented only by an unknown

number of participants in each condition (two studies from two records), the sample size in each

condition was instead estimated by dividing the total number of participants by the number of

conditions. Where means and standard deviations for the measure of intergroup bias were

available only for a change score representing the difference in intergroup bias pre- vs. post-

confrontation (one study from one record), we opted to extract data for the change score rather

than exclude the study. Where data were obtainable from a doctoral dissertation underpinning a

published record but not from the published record itself (one study from one record), we

included the doctoral dissertation and the data reported therein instead. We were unable to obtain

sufficient information to calculate effect sizes for three records (comprising four studies) and two

out of three intergroup bias measures reported in an additional record, so these were omitted

from the analysis. The final sample included in the analysis therefore consisted of 19 records,

comprising 32 separate studies and 91 effect sizes.

Extraction of Publication Information and Coding of Potential Moderators

Publication information and moderator variables were independently extracted and coded

by the first and third authors, using a standard coding protocol as detailed below. Author's

laboratory was coded by the first author only, given greater knowledge of key authors in the

field. A spreadsheet containing all extracted publication information, statistical data, and

moderator coding for each study is openly available from the Open Science Framework at

https://doi.org/10.17605/OSF.IO/2X3BY. The full dataset is also included in online

Supplemental Materials A.

Interrater reliability was measured using Cohen's kappa for categorical moderators

(interpreted in line with guidelines from Landis and Koch [1977] who suggested that $\kappa < .00 =$

poor; .00 - .20 = slight; .21 - .40 = fair; .41 -.60 = moderate; .61 - .80 = substantial; .81 - 1.00 =

almost perfect), and intraclass correlations (*ICC*) for continuous moderators (interpreted in line

with guidelines from Cicchetti [1994] who suggested that ICC < .40 = poor; .40 - .59 = fair; .60 -

.74 = good; .75 - 1.00 = excellent). Interrater reliability was substantial to almost perfect or

excellent in most cases (individual results reported below) and discrepancies were resolved

through discussion.

Characteristics of the Confrontation. The following characteristics of the confrontation

manipulation were coded: The group membership of the confronter relative to the participant

(ingroup; outgroup; irrelevant; mixed or unspecified: κ = 1), identity of the confronter (actual

target of intergroup bias; same group as target of intergroup bias; different group from target of

intergroup bias; computer; mixed or unspecified groups: κ = .68), target of confrontation

(experienced; observed; mixed confrontation: κ = .26), target of intergroup bias (Black people;

other ethnic minority groups; women; LGBTQ+ people; religious groups; people with

obesity/overweight; other/mixed targets: κ = 1), and whether the confrontation was performed in

public or private (public – witnessed by people other than the confronter; private – not witnessed

by people other than the confronter: κ = .69). Interrater reliability in all cases was either

substantial or almost perfect, except target of confrontation which had only fair interrater

reliability. In this case, discrepancies in coding were primarily driven by an error in coding of

one study that contributed a large number of effect sizes.

One additional moderator that was not pre-registered was also coded: Confrontation style.

This moderator was initially coded following the confrontation approaches identified in Chaney

and Sanchez's (2021) Prejudice Confrontation Styles Scale: educational, argumentative, help-

seeking, empathy, and humor, with the addition of an 'other' category. When reviewing this

coding strategy, it became apparent that most confrontations in the literature included an

educational component. Educational confrontations point out the instance of intergroup bias or

its negative impacts in a factual way, often labeling the instance as sexist or racist (e.g.,

"Statements like 'girls are dumb at math' are not really appropriate for this discussion.

Generalizations about what 'all men' or 'all women' are like are sexist. So let's be more careful in

how we describe the sexes": Boysen, 2013, p. 301; "You know, some of your responses seemed

a bit racist…the Black guy wandering the streets could be a lost tourist and this last guy could

work for the government. That's kind of racist, don't you think?": Simon et al., 2012, p. 4) and/or

pointing out the negative impact of intergroup bias on the target of bias (e.g.,

"...negative/positive stereotypes can create inequality": Burns & Granz, 2021, p. 509; "Hey man,

just remember that there are real people who arehurt *[sic]* when you harass them with that kind

of language": Munger, 2016, p. 636) or the person confronted (e.g., "...people might get upset at

you if you assumed they were a certain way because of their race": Burns, 2017, p. 49). Where

the other confrontation styles identified by Chaney and Sanchez were present, they were

predominantly used in addition to an educational component. Accordingly, confrontation style

was recoded using the following revised coding strategy. Confrontations were coded as

educational if they pointed out the instance of intergroup bias or its negative impacts, without

including any other style of confrontation. The remaining confrontations were coded as

argumentative, empathy, humor, help-seeking, or other, regardless of whether they also included

an educational component. That is, confrontations were coded as argumentative if the

confrontation included content that was aggressive or combative (e.g., "People are not going to

like you, they may not hire you...": Burns & Monteith, 2019, p.934; "I don't see why it's such a

damn problem. I am gay and had a straight roommate in the past. Anyone who has a problem

with it, should get over it": Hyers, 1999, p. 89), empathy when the confronter expressed their

own sadness or hurt resulting from the instance of intergroup bias ("It bothered me used how

much you used pretty common stereotypes about [muslims]": Kroeper, 2020, p. 56), and humor

if joking or sarcasm was used. Confrontations were coded as 'other' if they employed a

confrontation style or styles that did not fall under the above categories, for example,

confrontations that showed empathy towards the person confronted (e.g., "Such racial biases are

unfortunate among people who strive for social justice. You can choose to think about Blacks in

more open-minded ways": Burns & Monteith, 2019, p. 934; "... I know I do that without thinking

about it sometimes too, and that you already know this anyway. I'm sure you didn't mean to":

Lewis & Yoshimura, 2017, p.6), and confrontations that included a perspective-taking

component (e.g., "You likely wish to be viewed as a unique individual, and to have your

individuality acknowledged. Women, and all people, want this too": Burns, 2017, pp. 31-32).

The help-seeking category was included for completeness but was not expected to result in any

relevant data, given that this strategy does not involve a direct confrontation of prejudice, and

instead implicates seeking further guidance before confrontation or deferring confrontation to

another person. Coding of confrontation style had only fair interrater reliability ($\kappa = .39$).

Discrepancies were primarily due to disagreement over whether confrontations contained

components relating to empathy, as operationalized in the Prejudice Confrontation Styles Scale.

Discrepancies were resolved through discussion.

Characteristics of the Measure of Intergroup Bias. The following characteristics of the

measure of intergroup bias were coded: Type of intergroup bias measured ($\kappa = .77$) and whether

the measure was implicit or explicit ($\kappa = .49$). Type of intergroup bias was coded as affective

prejudice (e.g., feeling thermometer scores representing relative warmth towards Black vs. White

people: Simon et al., 2012), cognitive prejudice (representing measures that comprise belief-

based attitudes or 'isms' scales, e.g., modern sexism scale scores: Boysen, 2013; beliefs about

women in science: Parker et al., 2018), stereotyping (e.g., support for stereotypical career goals:

Burns, 2017; stereotype-consistent interpretations of person descriptions: Czopp et al., 2006),

behavioral intentions (e.g., intentions to monitor one's biases in the future: Burns & Granz, 2021;

Parker et al., 2018), behavior (e.g., self-reported avoidance of biased behaviour: Chaney &

Sanchez, 2018; use of racist slurs: Munger, 2016), or 'other': any other measure of intergroup

bias that did not fit into the above classifications (e.g., interest in gender-biased activities: Mark,

2007). Interrater reliability was only moderate for whether a measure was implicit or explicit, but

this was driven by errors in coding for only two effect sizes.

Characteristics of the Study Design. The following characteristics of the study design were

coded: Confrontation modality (verbal, written, mixed: $\kappa = .70$), study type (laboratory, field,

online, mixed: $\kappa = .94$), and time between confrontation and measurement of intergroup bias (in

days: $ICC = 1$). Time between confrontation and measurement was coded as 0 where the

confrontation and measurement of intergroup bias occurred in the same testing session or on the

same day. Where the time between the confrontation and measurement varied (e.g., 5-7 days:

Monteith et al., 2021), the lower time was entered to avoid overestimating the longevity of the

effects of confrontation.

Characteristics of the Sample. The following characteristics of the sample were extracted:

Overall sample size included in the analysis ($ICC = 1$), mean age ($ICC = 1$), gender composition

(% female: $ICC = .999$), ethnicity composition (% White: $ICC = 1$), sexual orientation

composition (% heterosexual: $ICC = 1$), country where participant sample was recruited ($\kappa = 1$),

and sample type (general population; student only sample; mixed student/general population

sample: $\kappa = .96$). In around half of the included studies, the sample relevant to the meta-analysis

was a subset of a larger sample (i.e., that included conditions that did not meet the meta-analysis

inclusion criteria) for whom demographic characteristics were reported. To reduce the burden on

the authors of these studies from data requests, we made a pragmatic choice to extract the

demographic data for the full sample of participants reported in the paper. All studies

randomized participants to conditions, and as such there are unlikely to be any systematic

differences in demographics between the full participant sample and subsample. Nonetheless,

any effects of demographics should be interpreted bearing this in mind.

Characteristics of the Publication. For published research, the year of publication was

extracted directly from the databases/platforms as part of the literature search. For unpublished

research, the year in which the research was completed was provided by the authors or extracted

from written materials (e.g., PhD thesis submission date). Publication status was categorized as

published or not published ($\kappa = 1$). Records that were categorized as not published when data

were extracted retained that classification even if later published (though the citation and year of

publication were updated as relevant). Author's laboratory was coded by the first author using

information from PhD theses, publicly available information on university and personal

academic websites, and prior knowledge of academic ancestry. The language of publication was

also coded ($\kappa = 1$), but all papers or summaries of research included were in English.

Risk of Bias Assessment

Risk of bias was assessed using the revised Cochrane risk-of-bias tool for randomized

trials (RoB 2: Sterne et al., 2019) which evaluates bias across five domains: (i) bias arising from

the randomization process, (ii) bias due to deviations from intended interventions, (iii) bias due

to missing outcome data, (iv) bias in measurement of the outcome, and (v) bias in selection of

the reported results) using five response options (no, probably no, no information, probably yes,

yes), which lead to overall judgments of 'low', 'some', or 'high' risk of bias for each of the five

domains, via set algorithms. All records included in the meta-analysis were independently

assessed for risk of bias by the first and third authors. Interrater reliability for the overall

judgment for each domain was almost perfect for bias arising from the randomization process (κ

= 1), bias due to deviations from intended interventions (κ = 1), bias due to missing outcome data

(κ = 1), bias in measurement of the outcome (κ = .89), bias in selection of the reported results (κ

= 1), and the overall risk of bias judgment (κ = 1). Discrepancies were resolved through

discussion.

Analysis Strategy

The standardized mean difference (*SMD*) was used as the measure of effect size, with

Hedges' *g* correction to mitigate small sample bias (Borenstein, 2009). Positive effect sizes

indicated that confronting prejudice reduced intergroup bias. Hedges' *g* and SE_g were calculated

by the first author using Comprehensive Meta-Analysis software, Version 3.3.070 (Borenstein et

al., 2013). As per Assink and Wibbelink (2016), the variance was calculated as SE^2. Hedges' *g*

was interpreted following Cohen's (1992) guidelines for interpretation of *d*, such that effect sizes

of 0.20, 0.50, and 0.80 were deemed to signify small, medium, and large effects, respectively.

Most studies included in the meta-analysis reported multiple eligible effect sizes for

overlapping participant samples (i.e., the 91 effect sizes were drawn from 36 independent

participant samples). To account for the statistical interdependency of effect sizes, a three-level

meta-analysis was fitted to report an overall pooled effect of confronting prejudice on intergroup

bias (e.g., Assink & Wibbelink, 2016; Viechtbauer, 2010). Traditional two-level meta-analyses

consider only two levels of variance in effect sizes: between participants variance (i.e., sampling

variance) and between studies variance. In addition to considering these levels of variance, three-

level meta-analyses also consider variance between effect sizes drawn from the same study

(López-López et al., 2018), analyzing the data according to this nested structure and capturing the variability in effect sizes at each of the three levels. That is, three-level meta-analyses model how effect sizes vary between participants in individual studies (level 1), between effect sizes within each study (level 2), and between studies (level 3; Assink & Wibbelink, 2016). This 'multi-level' analysis approach is recommended for effect sizes derived from the same participants using either separate measures or multiple time points (Assink & Wibbelink, 2016; Musekiwa et al., 2016).

The meta-analysis was conducted by the second and last authors using the rma.mv function in the *Metafor* package (Version 3.4-0; Viechtbauer, 2010) in the statistical software environment R (Version 4.3.2 'Eye Holes'; R Core Team, 2022). A mixed-effects model was fitted, with two grouping variables (level 2: effect size nested within level 3: individual studies) to define the nested structure using estimation based on the restricted maximum likelihood estimator (REML). The analysis produced a pooled effect size estimate across all studies and examined the variability in effect sizes across levels 1, 2, and 3 for the overall effect (the effect of confronting prejudice on intergroup bias) and the effects of the moderators. The data file and code used for analysis can be found at https://doi.org/10.17605/OSF.IO/2X3BY.

The methods for evaluating heterogeneity described in the pre-registered protocol (Cochran's Q statistic: Cochran, 1950; I^2 statistic: Higgins & Thompson, 2002) were replaced with alternative methods to better suit the three-level meta-analysis that was conducted. Specifically, the heterogeneity of effect sizes was examined by estimating the amount of variability in effect sizes at each of the three levels of analysis, using the formulas reported by Cheung (2014). Following recommendations by Assink and Wibbelink (2016), moderator analyses were considered justified if less than 75% of the variance could be attributed to between

participants variance (i.e., level-1 sampling variance). Cochran's Q-test for residual

heterogeneity and level-2 and -3 random-effects variance estimates (τ^2) are also reported for each

of the separate moderator analyses. Cochran's Q-test examines whether significant unexplained

variance between effect sizes remains after taking into account the variance explained by the

relevant moderator.

For categorical moderators, any levels of a moderator that was represented by fewer than

five effect sizes were excluded from the analysis (e.g., see Deeks et al., 2021; Fu et al., 2011).

For the moderator analyses, univariate subgroup analyses were conducted examining whether

variability in between-study effect sizes could be explained by each individual moderator (Hox,

2010). Specifically, each category within each moderator was dummy-coded (0: absent; 1:

present; Assink & Wibbelink, 2016), and all but one dummy-coded category (used as the

reference category) was entered into the moderation analysis; mean effects (and associated

standard errors, confidence intervals, and prediction intervals) for the omitted reference category

were then obtained through examining the 'intercept' variable (Assink & Wibbelink, 2016). This

process was then repeated with a different category as the reference category (i.e., omitting each

category in turn, and including the remaining ones) until mean effects for all categories had been

obtained. Cramér's V is reported between all possible pairs of categorical moderators to explore

associations between moderators. Hedges' g is reported for categorical moderators and

standardized betas (β) are reported for continuous moderators. 95% confidence intervals and

prediction intervals were used to determine whether the effect sizes for each category were

significantly different from zero. Where the overall test of moderation indicated a significant and

meaningful effect of the moderator (i.e., that mean effects for some of the categories differ in an

interpretable way), we also report an analysis comparing the category with the largest effect size to the remaining categories.

As outlined above, studies were graded on risk of bias in five areas, leading to a final qualitative evaluation of 'Low', 'Some Concern', or 'High' risk of bias. These were converted to numerical values as follows: 'Low' = 1; 'Some Concern' = 2; 'High' = 3, such that a higher score indicated greater concern regarding risk of bias. We assessed the potential impact of risk of bias on the overall effect size by entering risk of bias score as a categorical moderator in the main effects model.

Publication bias

Publication bias was examined through a moderator analysis comparing effect sizes in published vs. unpublished studies, and by using multi-level Egger's regression, as recommended when effect sizes are dependent (Rodgers & Pustejovsky, 2021). The pre-registered protocol initially indicated that we would also use Duval and Tweedie's (2000) Trim and Fill Technique to estimate the extent of publication bias, but this technique is not compatible with a three-level meta-analysis (Viechtbauer, 2017, October 31). A contour-enhanced funnel plot that plots each effect size (Hedges' *g*) against associated standard errors is included to provide a visual representation of potential publication bias. Shaded contours of different levels of statistical significance have been added to aid interpretation about causes of asymmetry, with publication bias indicated if studies are missing from low significance contours. Asymmetry resulting from studies missing from high significance contours suggests other systematic between-study differences are influencing publication (e.g., study quality).

Results

Descriptive Summary of Publication and Sample Characteristics

The records were published (or the data were collected, in the case of unpublished datasets) between 1999 and 2023 (*Mdn* = 2018). The median sample size was 87 participants (*SD* = 145.45, range = 43 to 869). The median age of the samples was 19.38 years (*SD* = 8.97 years, range = 18.71 to 41.55 years; missing data = 7) and the median percentage of women was 57.30% (*SD* = 20.32, range = 0 to 100%). Participants were predominantly White (*Mdn* = 97%, *SD* = 10.22, range = 68.60-100%; missing data = 1), and heterosexual (*Mdn* = 91.30%, *SD* = 4.84, range = 84.90-100%; missing data = 26). The majority of samples (63%) were students and every sample but one recruited participants from within the United States.

Main and Heterogeneity Analyses

The meta-analysis was conducted on 91 effect sizes from 36 independent samples. Analysis using a mixed-effects model indicated that, on average, confronting prejudice has a significant medium-sized effect on reducing intergroup bias, $g_+ = 0.54$, *SE* = 0.07, 95% *CI* [0.40, 0.67], 95% *PI* [-0.27, 1.34], $p < .001$, $Q = 423.75$, between-cluster variance (level 3) $\tau^2 = 0.12$, within-cluster variance (level 2) $\tau^2 = 0.04$ (See Figure 2). The distribution of variance in effect sizes at each level of analysis was as follows: level 1 = 17.67%; level 2: 18.41%; level 3: 63.92%. Following the recommendations of Assink and Wibbelink (2016), moderator analyses are considered justified because between-participants variance (level 1) is less than 75% of the total variance.

Publication Bias

Multi-level Egger's regression (Egger et al., 1997) indicated significant asymmetry in the funnel plot ($B = 4.16$ [1, 89], $p < .001$), such that more precise studies with larger positive effect sizes appeared to be missing from the distribution, along with less precise studies with smaller or negative effect sizes. In general, the patterns suggested that more precise studies tended to have

smaller effect sizes. While the degree of heterogeneity in effect sizes makes small-studies effects more difficult to interpret, examination of the distribution of effect sizes in the contour-enhanced funnel plot (see Figure 2) indicated that publication bias is unlikely to fully account for the asymmetry observed. Consistent with publication bias, there was an indication that some studies were missing from contour areas of low statistical significance. In contrast, inconsistent with publication bias, there are also likely missing studies in contour areas of high significance.

Individual Moderator Analyses

Univariate moderator analyses were conducted to examine the effect of characteristics of the confrontation, measure of intergroup bias, sample, and study design on the effect size. We also examined the impact of publication year and status, author's laboratory, and risk of bias score on the effect size. Five of the pre-registered moderators (target of confrontation; use of implicit vs. explicit measures of intergroup bias; country of participant recruitment; participant sexual orientation; time between confrontation and measurement of intergroup bias) were excluded from analysis due to a lack of data (only nine studies reported the sexual orientation of participants) or lack of variation (all but two studies examined experienced confrontation; only one study used an implicit measure of intergroup bias, all but one study recruited participants in the United States, and only six studies involved any type of intergroup bias measurement delay). Analyses of all other moderators are reported below. Results for all individual moderator analyses are reported in Table 1. Please note that where reported below, k refers to the number of effect sizes.

Characteristics of the Confrontation

The group membership of the confronter relative to participants, identity of the confronter, confrontation style, whether the confrontation was performed in public or private, or

target of intergroup bias did not significantly influence the effect of confronting prejudice on reducing intergroup bias. Effect sizes for each category of each moderator were all significantly different from zero (95% *CI*s not overlapping zero), except for identity of the confronter—unspecified/anonymous/mixed groups, confrontation style—empathy, public confrontations, and target of intergroup bias—other ethnic identities. For the analysis of the identity of the confronter, no effect sizes were coded as examining confrontation by the actual target of intergroup bias. For the analysis of confrontation style, no effect sizes were coded as examining humorous or help-seeking confrontation styles. For the analysis of the target of intergroup bias, there were fewer than five effect sizes coded as examining intergroup bias based on sexual orientation ($k = 2$), and other/mixed groups ($k = 3$), and no effect sizes coded as religion or weight-status. These categories are therefore not included in the relevant analyses.

Characteristics of the Intergroup Bias Measure

The type of intergroup bias measured significantly moderated the effect of confronting prejudice on reducing intergroup bias. Studies evaluating the effect of confrontation on stereotyping (which all measured stereotype use or endorsement) had a medium-to-large effect size that was significantly different from zero. Studies using measures of behavior or behavioral intentions (which all measured intentions to control or monitor future bias) had small-to-medium effect sizes that were significantly different from zero. Studies using measures of cognitive prejudice had a small effect size that was not significantly different from zero. There were fewer than five effect sizes coded as examining effects of confrontation on affective prejudice ($k = 2$), and miscellaneous other types of intergroup bias ($k = 2$). These categories are therefore not included in this analysis. When compared to studies evaluating the effect of confronting prejudice on stereotyping (the reference category), studies using measures of behavioral

intentions and cognitive prejudice had significantly smaller effect sizes. There was no significant difference in effect sizes between studies using measures of stereotyping and those using measures of behavior. These effects are illustrated in Figure 3.

Characteristics of the Sample

Age, gender, ethnicity or the type of sample did not significantly moderate the effect of confronting prejudice on reducing intergroup bias. Effect sizes for each category of the sample type moderator were significantly different from zero (95% *CI*s not overlapping zero).

Characteristics of the Study Design

Confrontation modality significantly moderated the effect of confronting prejudice on reducing intergroup bias. Studies employing verbal or written confrontations both had medium-to-large effect sizes that were significantly different from zero, whereas studies employing both modalities had a small effect size that was not significantly different from zero.

There was no moderating effect of study type on the effect of confronting prejudice on reducing intergroup bias. Effect sizes for laboratory and online studies were significantly different from zero (95% *CI*s not overlapping zero), whereas the effect size for field studies was not significantly different from zero. There were fewer than five effect sizes coded as mixed ($k = 1$), so this category was not included in this analysis.

Characteristics of the Publication

There were no significant effects of publication year, publication status, or author's laboratory on the effect of confronting prejudice on reducing intergroup bias. Effect sizes for each category of the publication status moderator and laboratory group moderator were significantly different from zero (95% *CI*s not overlapping zero).

Risk of Bias

There was no significant effect of overall risk of bias on the effect of confronting prejudice on reducing intergroup bias. Effect sizes for each category of this moderator were significantly different from zero (95% *CI*s not overlapping zero).

Associations between Moderators

Cramér's V between all pairs of categorical moderators is reported in Table 2. Author's laboratory and whether confrontation is performed in public or private showed significant, large associations with most other moderators. Identity of the confronter, target of intergroup bias, type of intergroup bias measured, confrontation modality, and study type had significant, large, or moderate to large associations with all or most other moderators. Group membership of the confronter relative to participants had significant, large associations with the identity of the confronter, target of intergroup bias, whether confrontation was performed in public or private, and author's laboratory, and significant, moderate to large associations with type of intergroup bias measured, confrontation modality, and study type. Confrontation style had significant, large associations with whether confrontation is performed in public or private, study type, and author's laboratory, and significant, moderate to large associations with the identity of the confronter, target of intergroup bias, type of intergroup bias measured, and sample type. Sample type had significant, large associations with the identity of the confronter and study type, and significant, moderate to large associations with the target of intergroup bias, confrontation style, type of intergroup bias measured, and confrontation modality. Publication status had significant, moderate to large associations with the target of intergroup bias, type of intergroup bias measured, and confrontation modality.

Discussion

The current meta-analysis aimed to estimate the effect of confronting prejudice on intergroup bias and evaluate factors that might influence the effect. A meta-analysis of 91 effect

sizes obtained from a systematic review demonstrated that, on average, confronting prejudice has

a significant, medium effect on reducing intergroup bias. This effect size is comparable to, or

slightly larger than, the small-to-medium effects found in contemporary meta-analyses of other

prejudice reduction interventions (i.e., Paluck et al., 2021) and suggests that confrontation of

prejudice is an effective strategy for reducing intergroup bias. There was only limited evidence

of publication bias. However, the samples in primary studies consisted mainly of young, White

adults recruited in the US, conferring limits on generalizability to other populations. Individual

analyses of moderators reflecting characteristics of the confrontation, measure of intergroup bias,

study design, sample, and publication, found two significant moderators. First, the type of

intergroup bias measured moderated the effect of confronting prejudice on intergroup bias, such

that confronting prejudice had a significant medium-to-large effect on the use and endorsement

of stereotypes, significant small-to-medium effects on behavior and behavioral intentions to

control or monitor future bias, and no significant effect on cognitive prejudice. Second,

confrontation modality moderated the effect of confronting prejudice on intergroup bias, such

that studies employing verbal or written confrontations had significant medium-to-large effect

sizes that were significantly different from zero, whereas studies employing both modalities in a

single confrontation had no significant effect on intergroup bias. The above findings and their

limitations are discussed further below, followed by suggestions for future research.

Evidence on Factors that Moderate the Effect of Confronting Prejudice on Intergroup Bias

The findings of the present review suggest that confronting prejudice has different effects

on different types of intergroup bias. Specifically, confronting prejudice had a significant

medium-to-large effect on the use and endorsement of stereotypes, followed by significant small-

to-medium effects on behavior and behavioral intentions to control or monitor future bias.

Conversely, there was no significant effect on cognitive prejudice, and there were insufficient

studies measuring affective prejudice to include in our analysis. When comparing different types

of intergroup bias, the effect of confronting prejudice on the use and endorsement of stereotypes

was significantly larger than the effects for all other measures of bias, except for behavioral

outcomes. These results suggest that confrontation is most effective at reducing the use and

endorsement of stereotypes, but ineffective at reducing prejudice, at least when operationalized

using common attitude or 'ism' scales that measure traditional and modern forms of cognitive

prejudice (e.g., Ambivalent Sexism Inventory: Glick & Fiske, 1996; Old-fashioned and Modern

Sexism scales: Swim et al., 1995).

This effect of confrontation on stereotype use and endorsement is clearly important—as

Czopp et al. (2006, p. 800) argues, even "situational compliance" to confrontation decreases the

burden on targets of intergroup bias and strengthens social norms denouncing bias. However, if a

reduction in cognitive prejudice is considered fundamental, there is a wide range of prejudice

reduction interventions that outperform confrontation, with significant (albeit often small) effects

on explicit attitudes (see Paluck et al., 2021 for a meta-analytic review). Similarly, although we

found that confronting prejudice increased behavioral intentions to control or monitor future

bias, we know from a vast body of literature on the intention-behavior gap (see Sheeran & Webb,

2016 for a review) that intentions do not always translate into action. A number of meta-analyses

have demonstrated that changes in intentions lead to much smaller changes in actual behavior

(e.g., Rhodes & Dickau, 2012; Webb & Sheeran, 2006). Although we found behavioral effects in

the current meta-analysis, this was based on a small number of effect sizes drawn from only two

studies (i.e., Chaney & Sanchez, 2018; Munger, 2016), making it difficult to draw strong

conclusions about whether people's intentions to control or monitor bias result in changes in

actual behavior. This issue is not limited to the literature on confronting prejudice—only 7% of

the outcomes included in Paluck et al.'s (2021) meta-analysis of prejudice reduction

interventions were behavioral, and criticisms leveled at social psychology in particular for

insufficient use of meaningful behavioral measures are well-known (e.g., see Baumeister et al.,

2007)

        We also note an important caveat to the interpretation of the effects of confrontation on

different types of intergroup bias. Studies measuring the effect of confronting prejudice on

stereotyping all involved private (vs. public) confrontation, performed in laboratory, online, or

mixed settings (vs. field settings), which tended to have larger effect sizes. It is therefore difficult

to disentangle the independent (or interactive) effects of each of these moderators. We return to

this issue in the limitations section. The research included in the current meta-analysis was also

dominated by a large cluster of studies (comprising over 40% of effect sizes) that used measures

of stereotype use (e.g., the classic sentence inference task first reported by Czopp et al., 2006) to

both confront and measure intergroup bias. It therefore remains unclear whether confrontation is

more effective at reducing the use of stereotypes (than other forms of intergroup bias), more

effective when confronting the use of stereotypes (than other forms of intergroup bias), or more

effective when there is high correspondence between the type of intergroup bias confronted and

measured.

        We faced similar challenges with a lack of variation in the literature when comparing the

effects of confronting prejudice in laboratory-based studies and their online equivalents relative

to more naturalistic field settings, as well as when examining the duration of effects. The

moderator analysis for study type was non-significant, implying that prejudice confrontation

works as well in the 'real world' as in the laboratory. However, the average effect size for field

studies was not significantly different from zero. In addition, all 8 effect sizes for field studies

came from one of the behavioral studies mentioned above—a novel study by Munger (2016) that

used bots to confront the use of racial slurs on Twitter, then measured subsequent use of these

slurs for up to 2 months. Accordingly, although we found little evidence that confronting

prejudice reduces intergroup bias in real-world settings, the effects for field studies were

inherently muddled with the potential decline of the effects of confronting prejudice over time.

In the current meta-analysis, we were unable to evaluate the persistence of the effects of

confrontation because most studies examined effects on intergroup bias immediately after

confrontation. Although there is some evidence that confrontation can continue to reduce

intergroup bias up to a week later (e.g., Burns & Monteith, 2019; Chaney & Sanchez, 2018;

Chaney et al., 2021), evaluation of longer-term effects is sparse and indicates that effects may

decay after 2 months (Munger, 2016). Others have noted this issue concerning the broader

literature evaluating prejudice reduction interventions (Hsieh et al., 2022; Paluck et al., 2021).

Notably, the overemphasis on the immediate impacts of confrontation on intergroup bias may

also not generalize well to real-world scenarios, which are likely to involve a delay between

confrontation and the next opportunity to avoid expressing intergroup bias.

Interestingly, the current meta-analysis found that confrontation modality moderated the

effect of confronting prejudice on intergroup bias. However, inspection of the patterns of effect

sizes indicated that moderation was driven by a non-significant effect of studies employing

mixed (i.e., both verbal and written) modalities. In addition, all effect sizes corresponding to

mixed modalities came from a multi-study paper where all studies used the same confrontation

paradigm. As such, this finding is not blatantly meaningful and is difficult to interpret.

The current meta-analysis otherwise largely found that the effects of confrontation are robust to differences in the characteristics of the confronter and how the confrontation is enacted. Notably, in line with the arguments by Wallace et al. (2024) that we discussed earlier, confrontations by both target and non-target (i.e., ally) groups appear to be equally effective at reducing intergroup bias. Given that allies generally experience less negative backlash to confrontation (e.g., Czopp & Monteith, 2003; Kutlaca et al., 2020) (Gulker et al., 2013), this suggests a particularly critical role for allies in taking responsibility for confronting prejudice. However, the current meta-analysis revealed that research has exclusively focused on the effects of (target or non-target group) *bystanders* confronting prejudice, as opposed to the target of the prejudice doing so. This focus is reflected in the terminology used in the broader literature— prejudice confrontation is also referred to as 'bystander anti-prejudice' (e.g., Stewart et al., 2014) and bystander anti-racism (e.g., Nelson et al., 2011). Similarly, Ashburn-Nardo and colleagues' Confronting Prejudiced Responses model (e.g., Ashburn-Nardo et al., 2008), which sets out the stages involved in deciding whether to confront intergroup bias, is based on Latané and Darley's (1970) seminal research on bystander intervention. Interestingly, this focus on action by bystanders may mirror what happens in reality. Recent research suggests that, at least in workplace settings, confrontations enacted by the person targeted by prejudice are less common than confrontations by non-target bystanders or allies (Dray & Sabat, 2022).

The confrontation style and whether the confrontation is performed in public do not appear critical to the effect of confronting prejudice on reducing intergroup bias. Although a relatively large number of studies have focused on identifying (Chaney & Sanchez, 2021) or comparing the impact of (e.g., Becker & Barreto, 2014; Czopp et al., 2006; Hyers, 2010; Martinez et al., 2017) different styles of confrontation, we found no significant effect of

confrontation style on reductions in intergroup bias, at least when operationalized in line with

Chaney and Sanchez's (2021) Prejudice Confrontation Styles scale. There was some evidence

that empathy-based confrontations may not be effective—the mean effect size for this

confrontation style was not significantly different from zero. However, most studies that used

empathy in confrontations were field studies, which were also characterized by a small non-

significant mean effect. As such, it is difficult to separate the effects of confrontation style from

those of the study type. In addition, as most confrontations included an educational component

(either exclusively or with another confrontation style), confrontations were necessarily coded

according to whether they used an educational style on their own, or instead employed

argumentative, empathy-based, or other styles *in addition* to an educational component. As such,

our findings indicate that confrontations that use argumentative or empathy-based styles

alongside educational styles are no more effective than confrontations that solely seek to educate.

These results are broadly consistent with previous literature demonstrating that the level of

hostility or threat in confrontations does not moderate their impact on intergroup bias (Czopp et

al., 2006; Hyers, 2010). Notably, no papers employed humorous confrontations, and thus the

current meta-analysis cannot shed light on whether these forms of confrontation are effective at

reducing intergroup bias. Although the current analysis indicated that confrontations that 'call

out' prejudice in public are no more or less effective than confrontations that 'call in' the

perpetrator in private, we also note that there were far fewer data points for public confrontations

(11 effect sizes from 3 studies) than private confrontations (80 effect sizes from 33 studies). As

such, further research is needed to explore the relative impact of public confrontation and its

nuances, for example, whether the nature or number of witnesses to public confrontations

matters.

It is worth noting that although the absence of moderation by target of intergroup bias is consistent with primary research studies that have found no evidence of differential effects of confrontation of gender vs. racial bias (Burns & Monteith, 2019; Parker et al., 2018), this finding is at odds with Monteith et al.'s (2022) arguments that confrontation of racism is more likely to be considered valid than confrontation of sexism (and thus in the context of the VIDOCC theoretical framework, more likely to promote a reduction in intergroup bias). An anonymous reviewer suggested that this may be due to the nature of confrontations employed by research that targets sexism – such that confrontations are often explicitly designed to try to persuade the recipient that their perpetration of sexism is real (e.g., by providing concrete evidence of bias: Parker et al., 2018), harmful (e.g., Burns & Granz, 2021), and at odds with prevailing social norms (e.g., Burns & Monteith, 2019). Although our meta-analysis includes data from studies or experimental conditions where these factors are absent as well as present, the inclusion of these manipulations may have attenuated differences in the effectiveness of confronting racism vs. sexism that would otherwise have been observed. To evaluate this explanation, future research should directly measure whether manipulating the above factors affects the perceived validity of confrontations and whether perceived validity then moderates or mediates subsequent effects on gender bias.

The current meta-analysis clearly demonstrates the lack of diversity in the targets of intergroup bias explored in research on confrontation. Studies tended to explore the effects of confronting prejudice against women (45% of effect sizes) and Black people (41% of effect sizes), neglecting other Minoritized identities. This issue is not exclusive to research on the effects of confronting prejudice. For example, although Paluck et al. (2021) excluded gender-based prejudice from their meta-analysis, they indicated that around a third of field-based

prejudice reduction research focused on prejudice towards African Americans. Other researchers have argued that our understanding of prejudice confrontation would benefit from an intersectional approach, considering target identity not as a singular construct, but as a multifaceted and interdependent set of identities that interact with participant characteristics and the broader sociocultural context (e.g., see Case et al., 2020; Remedios & Akhtar, 2019 for detailed discussions). For example, although the present review did not find that confrontation style moderated the effect of confronting prejudice on intergroup bias, we did not examine whether different styles of confrontation may be more or less effective when employed by different confronters (e.g., ingroup vs. outgroup; target vs. ally), or when directed at different groups of participants. Complex interactions may exist. For example, in view of the angry Black woman trope, it is plausible that argumentative confrontations enacted by Black women targeted by prejudice may not be as effective at reducing intergroup bias as argumentative confrontations enacted by allies. Black women are likely even more vulnerable than other targets to the 'backlash effect' (e.g., Stone et al., 2011) following bias confrontation, whereby confronted persons react extremely negatively to confrontation.

In the current meta-analysis, the effect of confronting prejudice on reducing intergroup bias seems largely robust to differences in participant sample characteristics. There was no evidence that the effects of confronting prejudice on reducing intergroup bias varied between samples as a function of age, gender composition, ethnicity composition, or type (general population, students). However, samples were relatively homogenous—although reasonably representative for gender, the included studies sampled from almost entirely U.S.-based and predominantly White and heterosexual populations and oversampled younger adults and student populations. As a result, it is unclear whether the findings of the current meta-analysis would

generalize to participant samples outside the U.S., older adults, and Minoritized groups. Meta-analytic research suggests, for example, that motivations for conformity are weaker in individualistic societies such as the U.S. Social norm explanations for the effects of confrontation would therefore predict stronger effects of confronting prejudice in countries with more collectivist values. Notably, research on the geographical distribution of prejudice within the U.S. indicates that our meta-analysis likely captured participant samples with relatively high levels of prejudice. 58 of the 91 effect sizes included in our meta-analysis came from participant samples recruited from single U.S. states (i.e., specific universities), namely Louisiana (Simon et al., 2012), Indiana (Burns, 2017; Burns & Granz, 2021; Burns & Monteith, 2019; Hildebrand et al., 2024; Kroeper, 2020; Monteith et al., 2021; Parker et al., 2018), Kentucky (Czopp et al., 2006; Mark, 2007), Montana (Lewis & Yoshimura, 2017), Pennsylvania (Hyers, 1999), New Jersey (Chaney & Sanchez, 2018; Chaney et al., 2021), and New York (Boysen, 2013). Research suggests that regional levels of prejudice in these states are high on at least one measure. Freng et al. (2022) for example, found that Indiana, Kentucky, New Jersey, and Pennsylvania scored above the median for pro-White bias on the IAT, and Kentucky, Louisiana, Montana, and New York scored above the median on the modern racism scale. Using Google search data, Stephens-Davidowitz (2014) found that Kentucky, Louisiana, and Pennsylvania ranked in the top 10 states for use of racial epithets in online searches. This suggests that confronting prejudice reduces intergroup bias even in populations with relatively high levels of prejudice – though we note that university student populations may not be representative of their broader state populations.

Concerning the effects of age, a recent meta-analysis by Hsieh et al. (2022) found that field-based prejudice reduction interventions were most effective in university/college-aged samples and least effective in non-student adults. This suggests that the effects of confrontation

may be attenuated in older samples, though the present review did not find that the age of the

sample moderated the effect of confronting prejudice on intergroup bias. It is also worth noting

that few studies in the present meta-analysis reported participants' sexual orientation.

Finally, in terms of characteristics of the publication, we found no evidence of a 'decline

effect'—publication year did not moderate the effect of confronting prejudice on intergroup bias.

At odds with other meta-analyses examining the effectiveness of prejudice reduction

interventions (e.g., Paluck et al., 2021), the current meta-analysis also found only limited

evidence of publication bias. That is, although the multi-level Egger's regression indicated

significant asymmetry in the funnel plot, this was driven not only by putative missing effect sizes

consistent with publication bias (missing smaller, non-significant effects) but also those

inconsistent with publication bias (missing larger, significant effects). The distribution of effect

sizes in the funnel plot also suggested that more precise studies tended to show smaller effects.

As such, although publication bias may account for some asymmetry in the funnel plot, patterns

may also be explained by variations in study quality between small and large studies. Further

research using larger sample sizes will help evaluate this explanation and better pinpoint the size

of the effect of confronting prejudice on reducing intergroup bias.

Limitations

In the current meta-analysis, we sought to include literature published in languages other

than English to avoid a mono-language bias (see Johnson, 2021). The databases and platforms

we used to search published literature (Scopus, Web of Science, PsycInfo) include records

written in languages other than English, and the preprint platform used to search grey literature

(OSF Preprints) samples from a range of geographically diverse preprint servers (and thus a

potentially varied range of languages). However, we acknowledge that English-language records

are overrepresented in the databases and platforms we used to conduct our searches and that our final dataset consists solely of studies reported in English. As such, our results may not generalize to studies reported in other languages.

We also note that although we tested a range of potential moderators of the effects of confronting prejudice on reducing intergroup bias, only a small number of moderators were statistically significant. Although the lack of moderation may reflect the relative stability of the effects of confronting prejudice, the moderator analyses may also be underpowered, given the modest size of the prejudice confrontation literature. In addition, there were large associations between a number of categorical moderators. Author's laboratory was strongly associated with almost all other categorical moderators, reflecting the tendency for laboratory members to operationalize and evaluate the effects of confrontation in familiar 'tried and tested' ways. If there is sufficient cross-laboratory variation in characteristics of the confrontation, measures of intergroup bias, study design, and sample, these associations are not inherently problematic, and are instead likely to facilitate research. Similarly, other associations reflect intrinsic or pragmatic relationships between moderators – online studies overwhelmingly confronted participants via computer-delivered text; the group membership of the confronter relative to participants was necessarily coded as 'irrelevant, mixed, or unspecified ' for computer-based confrontations; confrontations by ingroup members were entirely performed by people from a different group from the target of prejudice (i.e., there were no instances where participants were confronted for showing prejudice towards their ingroup); university student samples were most often associated with laboratory studies whereas general population samples participated online; and although public confrontations were seen in both laboratory and field studies, confrontations were always private in online studies or studies where confrontation was delivered via computer.

Conversely, other clusters of characteristics are not as easily explained, nor are their individual effects easily disentangled, leaving gaps in our understanding of which factors moderate the effect of confronting prejudice on intergroup bias. For example, relative to private confrontations, studies using public confrontations were more likely to involve outgroup confronters and empathy-based confrontations and measure the impact on behavior. In contrast, studies using private confrontations were more likely to involve ingroup confronters and educational confrontations and measure the impact on stereotyping, cognitive prejudice, or behavioral intentions. To highlight the most extreme gaps in coverage, we extracted 61 effect sizes evaluating educational confrontations performed in private whereas we extracted only 1 effect size evaluating educational confrontations performed in public. Similarly, the impact of empathy-based confrontation styles on intergroup bias has not yet been tested in online studies, whereas the impact of solely educational confrontation styles on intergroup bias has not been tested in the field. Finally, in studies where the confronter was not a member of the group targeted by bias (i.e., confrontation by allies), confrontations were most likely to target prejudice towards Black people, whereas computer-based confrontations were most likely to target gender-based prejudice – and we found no studies that tested whether confrontation by the actual target of bias (i.e., the woman towards whom bias was directed) reduces gender-based prejudice. Further primary research that systematically manipulates and evaluates the impact of these potential moderator variables in fully factorial designs is therefore needed.

Directions for Future Research

The above discussion identifies clear limitations and gaps in evidence that need to be addressed before the potential of confrontation as a prejudice reduction strategy can be realized. If we categorize the reviewed literature using Ijzerman et al.'s (2020) innovative Evidence

Readiness Level (ERL) taxonomy, it is clear that further laboratory-based research is needed to test theoretical explanations of the effects of confronting prejudice on intergroup bias and evaluate long-term effects of confrontation across different types of intergroup bias and diverse participant samples; alongside more field-based research that evaluates effects of confronting prejudice on meaningful outcomes in real-world settings. We set out our suggested priorities for future research below.

Further Testing of Theoretical Explanations of the Effects of Confronting Prejudice on Intergroup Bias

Further research is needed to better establish the causal mechanisms underlying the effects of confrontation on intergroup bias. Although there is evidence in the existing literature that negative self-directed affect and rumination (key mechanisms implicated in the VIDOCC theoretical framework: Monteith et al., 2022) mediate the effects of confronting prejudice on intergroup bias (Alt et al., 2019; Burns & Granz, 2021; Chaney & Sanchez, 2018; Parker et al., 2018), no research has directly measured the perceived validity of confrontations to examine whether regarding a confrontation as valid is a necessary precondition of self-regulation of prejudice. Similarly, research has only recently begun to measure whether confrontation strengthens perceived social norms regarding the expression of intergroup bias (see Li et al., 2024), and no research has examined whether changes in norm strength mediate the effects of confronting prejudice on intergroup bias. Further research that includes measures of the perceived validity of confrontations and social norms regarding the expression of intergroup bias would help develop our understanding of the mechanisms underlying the effect of confrontation on bias reduction.

These measures could also be used to explore participant characteristics that may predict whether one's own standards or the standards of others (i.e., social norms) are used as the basis for consideration of prejudice-related discrepancies. A better understanding of the interaction between participant characteristics and theoretical mechanisms would give researchers more guidance on which aspects of confrontation could be emphasized to enhance their effectiveness and in which populations. For example, although people often overestimate the degree to which their attitudes towards different social groups are shared by society as a whole (the 'false consensus effect')—this effect is stronger among more prejudiced people (Pederson et al., 2008; Watt & Larkin, 2010). Injunctive social norms are therefore particularly likely to be an important contributor to self-regulation of prejudice in high-prejudice people, given that unlike low-prejudice people, they may not experience discrepancies between self-derived standards and actual prejudice expressed. As Czopp et al. (2006, p. 785) states, "although high-prejudice people do not have such internalized standards for nonprejudice, they recognize the social stigma associated with being publicly labeled a bigot." Initial research suggests that confrontations are most effective at reducing sexism in high-prejudice people when they occur alongside exposure to anti-sexist social norms (Mark, 2007).

Given the consistent and sizeable effects of confrontation on the use of stereotypes shown in the current meta-analysis, researchers may want to use the tried-and-tested sentence inference paradigm (e.g., Burns & Granz, 2021; Burns & Monteith, 2019; Chaney & Sanchez, 2018; Czopp et al., 2006; Lewis & Yoshimura, 2017) when testing theoretical explanations, at least initially, in order to focus efforts on generating theoretical insight rather than on testing new paradigms.

Evaluation of Long-Term Effects of Confrontation Across Different Types of Intergroup Bias and Diverse Participant Samples

Further research is needed to evaluate the effects of different styles and sources of confrontation across a broader range of types of intergroup bias, in more diverse participant samples, and over longer time frames. Research that systematically manipulates the type of intergroup bias both confronted and measured is critical to establish whether effects of confrontation reliably extend to behavioral outcomes and measures of affective prejudice, and to disentangle effects of what type of intergroup bias is confronted from what type of intergroup bias is measured (and the correspondence between the two). Given the absence of research examining confrontation by individuals directly targeted by intergroup bias, research may seek to examine whether confrontation of prejudice by individual targets of prejudice has the same effect on reducing intergroup bias as confrontation of prejudice by bystanders, whilst bearing in mind the personal costs that may be involved in confronting prejudice directed against oneself. Similarly, future research should seek to expand our understanding of the consequences of confronting prejudice by considering a range of different and intersecting identities alongside different confrontation styles, including those that use humor. An additional avenue for future research is to evaluate the effects of observed and imagined confrontation of prejudice on intergroup bias. Although we planned to compare the effectiveness of experienced vs. observed confrontation in our meta-analysis, only one study manipulated observed confrontation, precluding moderator analysis. Similarly, although examining the effect of imagined confrontation on intergroup bias was beyond the scope of our pre-registration, again, our literature searches revealed only one relevant study (i.e., Munder et al., 2024). Given the effectiveness of other observation- and imagery-based prejudice reduction interventions (e.g., see

meta-analyses on vicarious intergroup contact: Banas et al., 2020; imagined intergroup contact:

Miles & Crisp, 2014), primary research should seek to evaluate whether observing and

imagining confrontation of prejudice also reduces intergroup bias.

Future research should also focus on increasing diversity in participant samples to

explore the generalizability of the effects of confronting prejudice on intergroup bias. There has

been widespread criticism of the homogeneity of participant samples in psychological research

(e.g., Arnett, 2008; Rad et al., 2018; Shen et al., 2011), and the literature on confronting

prejudice is no exception, primarily drawing on participant samples that are White, young, and

recruited from the U.S. Further research is therefore needed to evaluate whether confronting

prejudice reduces intergroup bias in a more diverse range of participants in settings outside the

U.S.

Research should also explore the degree to which the effects of confronting prejudice are

time sensitive. As noted earlier, the vast majority of studies in this field examine the effects of

confrontation on intergroup bias measured immediately after confrontation, and thus less is

known about the persistence of effects, particularly after periods over one week. In addition, we

note that all studies included in the current meta-analysis manipulated confrontation as an

immediate or timely single response to an incident of intergroup bias. We therefore do not know

whether delayed confrontations (e.g., approaches that later 'call in' the transgressor) are as

effective as an immediate response. This is critical given evidence that a substantial proportion

of confrontations occur after a delay (e.g., Dray & Sabat, 2022). We also do not know whether

multiple confrontations have a cumulative effect on prejudice reduction or are indeed necessary

for the persistence of effects. The model of self-regulation of prejudice (e.g., Monteith, 1993;

Monteith et al., 2002) encompassed in the VIDOCC theoretical framework (Monteith et al.,

2022) proposes that over time, the stimuli associated with expressions of bias and negative self-directed affect become cues for control, which trigger subsequent self-regulatory action to control prejudiced responses. What is unclear, is whether one instance of confrontation is sufficient to consolidate cues for control. Some of these issues may be best addressed in longitudinal field research, which brings us to our final priority for future research.

Evaluation of Effects of Confronting Prejudice on Meaningful Outcomes in Real-World Settings

Research should also move to evaluate whether confronting prejudice drives meaningful changes in intergroup bias in naturalistic settings that more closely reflect how confrontation would be employed in the real world. There is much to learn about potential avenues for field studies from current research on prejudice reduction. For example, in relation to strategies for manipulating confrontation, Paluck's (2011) research on the effects of peer influence on intergroup bias trained student leaders ("peer trainers") to confront prejudice in school settings, randomly allocating schools to an intervention condition where peer trainers were deployed in the autumn or a waitlist control where the student leaders were deployed in the spring. Other field studies exemplify strategies for measuring the real-world impacts of interventions. Mousa (2020) formed a sports league to evaluate the longitudinal effects of intergroup contact between Christians and Muslims in Iraq, examining the impact on a variety of contextually relevant behavioral outcomes such as whether Christian participants donated their study payment to an organization benefitting both Christians and Muslims (vs. one benefiting only Christians). As a guide, Paluck et al. (2021, p. 554) suggest selecting field sites "based on the kinds of behavioral outcomes that are possible to observe". Chang et al. (2019) for example, drew on data from a large global organization to examine the effect of diversity training on behavioral measures of
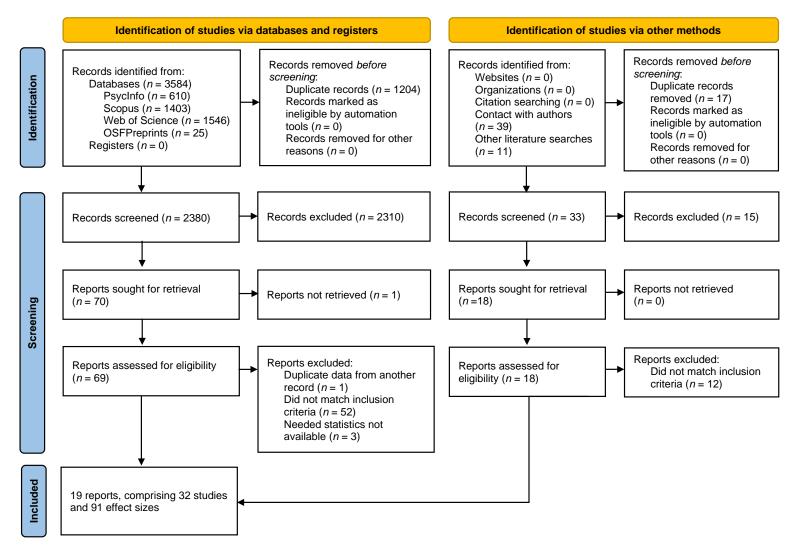
intergroup bias including the number of racially Minoritized people and women nominated for

internal awards for excellence, or selected as mentees. As highlighted earlier, Munger's (2016)

research on confronting prejudice instead harnessed the opportunities provided by social

networking sites. They identified Twitter accounts that tweeted racial slurs, used bot accounts to

tweet replies that confronted that behavior, and then measured subsequent use of racial slurs by

the perpetrating accounts over an extended period of time. Social media platforms may provide

particularly flexible opportunities to address many of the research gaps highlighted in the current

meta-analysis, including manipulating the identity of the confronter, the confrontation style, the

time frame of confrontation, and also measuring the duration of effects. Mosleh et al. (2022)

provide a useful primer and discussion of ethical considerations for field studies using social

media.

Conclusions

The findings of the current systematic review with meta-analysis indicate that in samples

predominantly made up of young, White, US-based adults, confronting prejudice reduces the

immediate use or endorsement of stereotypes and increases intentions to monitor and control

intergroup bias in the future, but does not change cognitive forms of prejudice. To advance our

understanding of the theoretical mechanisms that underlie the effects of confronting prejudice on

intergroup bias and better harness its potential as a prejudice reduction strategy, more direct tests

of theoretical mechanisms in primary studies are needed. Although current evidence suggests

that the effects of confronting prejudice are largely unaffected by who does the confronting and

how they do so, our understanding of the factors that moderate the effects of confronting

prejudice on intergroup bias, and how they might apply in real-world settings, is currently

incomplete. Further research is needed to evaluate the long-term effects of confronting prejudice

across different types of intergroup bias and more diverse participant samples, alongside research

that evaluates the effects of confronting prejudice on meaningful intergroup outcomes in real-

world settings. We therefore echo calls to diversify research on prejudice confrontation to aid

generalizability and better capture the complexity of intergroup bias and its mitigation.

Figure 1

PRISMA search flow diagram

Identification of studies via databases and registers

Records identified from:
- Databases (*n* = 3584)
    - PsycInfo (*n* = 610)
    - Scopus (*n* = 1403)
    - Web of Science (*n* = 1546)
    - OSFPreprints (*n* = 25)
- Registers (*n* = 0)

Records removed *before screening*:
- Duplicate records (*n* = 1204)
- Records marked as ineligible by automation tools (*n* = 0)
- Records removed for other reasons (*n* = 0)

Records screened (*n* = 2380)

Records excluded (*n* = 2310)

Reports sought for retrieval (*n* = 70)

Reports not retrieved (*n* = 1)

Reports assessed for eligibility (*n* = 69)

Reports excluded:
- Duplicate data from another record (*n* = 1)
- Did not match inclusion criteria (*n* = 52)
- Needed statistics not available (*n* = 3)

Identification of studies via other methods

Records identified from:
- Websites (*n* = 0)
- Organizations (*n* = 0)
- Citation searching (*n* = 0)
- Contact with authors (*n* = 39)
- Other literature searches (*n* = 11)

Records removed *before screening*:
- Duplicate records removed (*n* = 17)
- Records marked as ineligible by automation tools (*n* = 0)
- Records removed for other reasons (*n* = 0)

Records screened (*n* = 33)

Records excluded (*n* = 15)

Reports sought for retrieval (*n* =18)

Reports not retrieved (*n* = 0)

Reports assessed for eligibility (*n* = 18)

Reports excluded:
- Did not match inclusion criteria (*n* = 12)

19 reports, comprising 32 studies and 91 effect sizes

Identification

Screening

Included

Note. n = Number of records or reports

Figure 2

Contour-enhanced funnel plot of effect sizes

Note. Effect sizes (x-axis) are visualised with their related standard errors (y-axis). Effect sizes (Hedges' *g*) were coded such that positive effect sizes indicated that prejudice confrontation reduced intergroup bias. Study effect sizes are shown as circles. The overall mean effect is shown as a vertical line. Shaded contour lines depict levels of statistical significance (see figure legend).

Figure 3

Forest plot for type of intergroup bias measured

Table 1

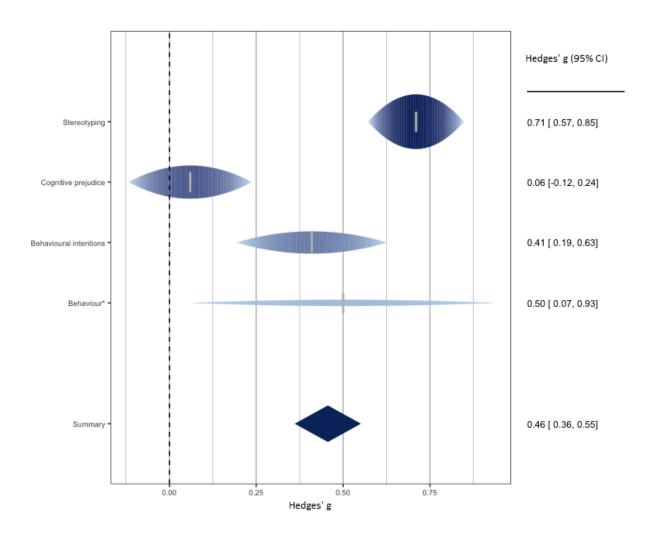*Summary of Moderator Analyses*

Variable	k	F(df)	β (SE)	p	g+(SE)	CI-, CI+	PI-, PI+	Q (p)	Residual τ^2
Characteristics of the confrontation									
Group membership of the confronter relative to participants	91	1.29(2, 88)		.28				420.01 (<.001)	L2: 0.04 L3: 0.12
Ingroup	19				0.69(0.12)	0.46, 0.92	-0.14, 1.52		
Outgroup	11				0.59(0.14)	0.30, 0.88	-0.26, 1.43		
Irrelevant/mixed/ unspecified	61				0.46(0.09)	0.28, 0.63	-0.36, 1.27		
Identity of the confronter	91	1.45(3, 87)		.24				404.62 (<.001)	L2: 0.04 L3: 0.12
Same group as target of intergroup bias	10				0.56(0.15)	0.27, 0.85	-0.27, 1.39		
Different group from target of intergroup bias	24				0.64(0.10)	0.45, 0.85	-0.16, 1.45		
Computer	38				0.55(0.11)	0.31, 0.78	-0.27, 1.36		
Mixed/unspecified group	19				0.26(0.16)	-0.06, 0.58	-0.58, 1.10		
Target of intergroup bias	91	1.89(4, 86)		.12				400.86 (<.001)	L2: 0.04 L3: 0.11
Black people	37				0.69(0.09)	0.51, 0.86	-0.09, 1.46		
Women	41				0.40(0.10)	0.19, 0.60	-0.38, 1.18		
Other ethnic identities	8				0.40(0.26)	-0.13, 0.93	-0.52, 1.32		
Confrontation style	91	0.76(3, 87)		.52				389.74(<.001)	L2: 0.04 L3: 0.13
Argumentative	7				0.43(0.16)	0.12, 0.74	-0.43, 1.29		
Educational only	62				0.60(0.08)	0.44, 0.76	-0.22, 1.42		
Empathy	10				0.35(0.27)	-0.18, 0.88	-0.62, 1.32		

	k	F(df)	p	Estimate(SE)	95% CI	PI	Q (p)	Variance
Other	12			0.44(0.13)	0.19, 0.69	-0.40, 1.28		
Whether confrontation is performed in public or private	91	1.06(1, 89)	.30				422.73 (<.001)	L2: 0.04 L3: 0.13
Public	11			0.30(0.24)	-0.18, 0.78	-0.63, 1.23		
Private	80			0.56(0.07)	0.42, 0.70	-0.25, 1.37		
Characteristics of the measure of intergroup bias								
Type of intergroup bias measured	91	9.04(5, 85)	<.001				254.57 (<.001)	L2: 0.02 L3: 0.08
Cognitive prejudice	22			0.06(0.09)	-0.12, 0.24	-0.57, 0.69		
Stereotyping	45			0.71(0.07)	0.58, 0.84	0.09, 1.33		
Behavior	10			0.50(0.22)	0.07, 0.93	-0.24, 1.25		
Behavioral intentions	10			0.41(0.11)	0.19, 0.63	-0.23, 1.05		
Characteristics of the study design								
Confrontation modality	91	3.16(2, 88)	.047				380.93 (<.001)	L2: 0.04 L3: 0.10
Verbal	15			0.64(0.13)	0.39, 0.89	-0.15, 1.43		
Written	66			0.56(0.08)	0.40, 0.71	-0.20, 1.32		
Mixed	10			0.05(0.21)	-0.36, 0.46	-0.80, 0.90		
Study type	91	0.43(3, 87)	.73				394.85 (<.001)	L2: 0.04 L3: 0.14
Laboratory	47			0.58(0.09)	0.40, 0.75	-0.26, 1.42		
Field	8			0.31(0.38)	-0.44, 1.07	-0.80, 1.43		
Online	35			0.46(0.13)	0.21, 0.72	-0.40, 1.33		
Characteristics of the sample								
Mean age	72	-.001(.008)	.95				319.39 (<.001)	L2: 0.04 L3: 0.11
Gender composition (% women)	91	.004(.003)	.20				421.48 (<.001)	L2: 0.04 L3: 0.12
Ethnicity composition (% White)	89	.01(.01)	.17				423.29 (<.001)	L2: 0.04 L3: 0.13
Sample type	91	0.20(1, 89)	.66				410.08 (<.001)	L2: 0.04 L3: 0.13
General population	33			0.49(0.13)	0.22, 0.75	-0.36, 1.34		

	k	F(df)	β(SE)	p	g+(SE)	CI-, CI+	PI-, PI+	Q(p)	Residual τ²
Students	58				0.56(0.08)	0.40, 0.72	-0.27, 1.38		
Characteristics of the publication									
Publication year	91		0.005(0.01)	.70				422.96 (<.001)	L2: 0.04 L3: 0.13
Publication status	91	1.74(1, 89)		.19				402.99 (<.001)	L2: 0.04 L3: 0.12
Published	53				0.62(0.09)	0.44, 0.80	-0.19, 1.43		
Not published	36				0.43(0.10)	0.24, 0.64	-0.38, 1.25		
Author's Laboratory	91	2.00(2, 88)		.14				387.39 (<.001)	L2: 0.04 L3: 0.12
Monteith	63				0.53(0.09)	0.36, 0.70	-0.26. 1.32		
Chaney	9				0.79(0.16)	0.47, 1.11	-0.05, 1.63		
Other	19				0.36(0.14)	0.08, 0.64	-0.46, 1.18		
Risk of bias									
Overall risk of bias	91	3.49(1, 89)		.07				375.59 (<.001)	L2: 0.04 L3: 0.11
High	53				0.64(0.09)	0.47, 0.82	-0.14, 1.42		
Some concern	38				0.40(0.10)	0.20, 0.59	-0.39, 1.18		

Note. k = Number of effect sizes; $F(df)$ = F-test for moderator analysis of categorical variables that tests whether subgroup mean effect sizes are significantly different from each other (degrees of freedom); $\beta(SE)$ = Standardised beta (standard error) for moderator analysis of continuous variables; p = p-value for F-test or β ($p<.05$ indicates significant differences in subgroup mean effect sizes, or a significant relationship between the moderator and effect sizes); $g_+(SE)$ = standardized mean difference effect size with Hedges' adjustment (standard error), coded such that positive effect sizes indicated that prejudice confrontation reduced intergroup bias; *CI-, CI+* = upper and lower 95% confidence intervals; *PI-, PI+* = upper and lower 95% prediction intervals; $Q(p)$= Cochran's Q-test for residual unexplained heterogeneity after testing whether the moderator variable explains any variability (p-value; $p<.05$ indicates statistically significant variability in effect sizes is still present); Residual τ^2 = Tau-squared variance estimate for between-clusters (L3 = level 3) and within-clusters (L2 = level 2) after adding the moderator.

Table 2

Association between categorical moderators measured by Cramér's V

Moderator	1	2	3	4	5	6	7	8	9	10	11
1. Group membership of the confronter relative to participants	-										
2. Identity of the confronter	.91***	-									
3. Target of intergroup bias	.54***	.58***	-								
4. Confrontation style	.27*	.31**	.36***	-							
5. Whether confrontation is performed in public or private	.64***	.58***	.50***	.74***	-						
6. Type of intergroup bias measured	.45***	.37***	.41***	.49***	.75***	-					
7. Confrontation modality	.46***	.62***	.38***	.22	.13	.40**	-				
8. Study type	.44***	.57***	.42***	.51***	.84***	.53***	.42***	-			
9. Sample type	.11	.53***	.33*	.37**	.28**	.37*	.46***	.81***	-		
10. Publication status	.11	.07	.38*	.19	.18	.44**	.35**	.29	.13	-	
11. Author Laboratory	.60***	.60***	.51***	.52***	.72***	.48***	.56***	.51***	.26*	.02	-

Note. $*p < .05. **p < .01. ***p < .001.$

References

Asterisked citations included at least one study that is included in the meta-analysis.

Abrams, D., & Hogg, M. A. (1990). Social identification, self-categorization and social influence. *European Review of Social Psychology*, *1*(1), 195-228. https://doi.org/10.1080/14792779108401862

Abrams, D., Swift, H., & Houston, D. (2018). Developing a national barometer of prejudice and discrimination in Britain. *Equality and Human Rights Commission*, *Research Report 119*. https://www.equalityhumanrights.com/sites/default/files/national-barometer-of-prejudice-and-discrimination-in-britain.pdf

Alt, N. P., Chaney, K. E., & Shih, M. J. (2019). "But that was meant to be a compliment!": Evaluative costs of confronting positive racial stereotypes. *Group Processes & Intergroup Relations*, *22*(5), 655-672. https://doi.org/10.1177/1368430218756493

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3-25. https://doi.org/10.1037/amp0000191

Arnett, J. J. (2008). The neglected 95: Why American psychology needs to become Less American. *American Psychologist*, *63*(7), 602-614. https://doi.org/10.1037/0003-066X.63.7.602

Ashburn-Nardo, L., Morris, K. A., & Goodwin, S. A. (2008). The confronting prejudiced responses (CPR) model: Applying CPR in organizations. *Academy of Management Learning & Education*, *7*(3), 332-342. https://doi.org/10.5465/amle.2008.34251671

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, *12*(3), 154-174. https://doi.org/http://dx.doi.org/10.20982/tqmp.12.3.p154

Banas, J. A., Bessarabova, E., & Massey, Z. B. (2020). Meta-analysis on mediated contact and prejudice. *Human Communication Research*, *46*(2-3), 120-160. https://doi.org/10.1093/hcr/hqaa004

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396-403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Becker, J. C., & Barreto, M. (2014). Ways to go: Men's and women's support for aggressive and nonaggressive confrontation of sexism as a function of gender identification. *Journal of Social Issues*, *70*(4), 668-686. https://doi.org/10.1111/josi.12085

Becker, J. C., Zawadzki, M. J., & Shields, S. A. (2014). Confronting and reducing sexism: A call for eesearch on intervention. *Journal of Social Issues*, *70*(4), 603-614. https://doi.org/10.1111/josi.12081

Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, *79*(6), 993-997. https://doi.org/10.1037/0021-9010.79.6.993

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2013). Comprehensive Meta-Analysis, Version 3.3.070

Borenstein, M. H., L. V.; Higgins, J. P.; Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley.

*Boysen, G. A. (2013). Confronting math stereotypes in the classroom: Its effect on female college students' sexism and perceptions of confronters. *Sex Roles*, *69*(5-6), 297-307. https://doi.org/10.1007/s11199-013-0287-y

*Burns, M. D. (2017). *How to confront a compliment? Reactions to confrontations of positive stereotype use.* [Doctoral Dissertation, Purdue University]. Purdue e-Pubs. https://docs.lib.purdue.edu/dissertations/AAI10601207/

*Burns, M. D., & Granz, E. L. (2021). Confronting sexism: Promoting confrontation acceptance and reducing stereotyping through stereotype framing. *Sex Roles*, *84*(9-10), 503-521. https://doi.org/10.1007/s11199-020-01183-5

*Burns, M. D., & Monteith, M. J. (2019). Confronting stereotypic biases: Does internal versus external motivational framing matter? *Group Processes & Intergroup Relations*, *22*(7), 930-946. https://doi.org/10.1177/1368430218798041

Case, K. A., Rios, D., Lucas, A., Braun, K., & Enriquez, C. (2020). Intersectional patterns of prejudice confrontation by White, heterosexual, and cisgender allies. *Journal of Social Issues*, *76*(4), 899-920. https://doi.org/10.1111/josi.12408

*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. https://doi.org/10.1177/0146167217741344

*Chaney, K. E., & Sanchez, D. T. (2019). *The endurance of interpersonal confrontations as a prejudice reduction strategy* [Data set]. Open Science Framework. http://osf.io/bj9hv

Chaney, K. E., & Sanchez, D. T. (2021). Prejudice confrontation styles: A validated and reliable measure of how people confront prejudice. *Group Processes & Intergroup Relations*, 136843022110058. https://doi.org/10.1177/13684302211005841

*Chaney, K. E., Sanchez, D. T., Alt, N. P., & Shih, M. J. (2021). The breadth of confrontations as a prejudice reduction strategy. *Social Psychological and Personality Science*, *12*(3), 314-322. https://doi.org/10.1177/1948550620919318

Chaney, K. E., Sanchez, D. T., & Remedios, J. D. (2022). Confronting first impressions: Motivating self-regulation of stereotypes and prejudice through prejudice confrontation. In E. Balcetis & G. Moskowitz (Eds.), *The Handbook of Impression Formation* (1st ed., pp. 435-458). Routledge.

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), 363-371. https://doi.org/https://doi.org/10.1037/tps0000042

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. *Proceedings of the National Academy of Sciences*, *116*(16), 7778-7783. https://doi.org/10.1073/pnas.1816076116

Cheung, M. W. L. (2014). Modeling dependent effect sizes with rhree-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, *19*(2), 211-229. https://doi.org/10.1037/a0032968

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284-290. https://doi.org/10.1037/1040-3590.6.4.284

Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, *37*(3/4), 256. https://doi.org/10.2307/2332378

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. https://doi.org/https://doi.org/10.1037/0033-2909.112.1.155

Cowan, G., & Hodge, C. (1996). Judgments of hate speech: The effects of target group, publicness, and behavioral responses of the target. *Journal of Applied Social Psychology*, *26*(4), 355-374. https://doi.org/10.1111/j.1559-1816.1996.tb01854.x

Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, *82*(3), 359-378. https://doi.org/10.1037/0022-3514.82.3.359

Czopp, A. M., & Monteith, M. J. (2003). Confronting prejudice (literally): Reactions to confrontations of racial and gender bias. *Personality and Social Psychology Bulletin*, *29*(4), 532-544. https://doi.org/10.1177/0146167202250923

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

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2021). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)*. Cochrane. Available from www.training.cochrane.org/handbook

Dray, K. K., & Sabat, I. E. (2022). Confronting sexism: Identifying dimensions and exploring impact. *Journal of Applied Social Psychology*, *52*(5), 316-340. https://doi.org/10.1111/jasp.12861

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

Eliezer, D., & Major, B. (2012). It's not your fault: The social costs of claiming discrimination on behalf of someone else. *Group Processes & Intergroup Relations*, *15*(4), 487–502. https://doi.org/10.1177/1368430211432894

Freng, S., Schweitzer, K., Estrada-Reynolds, V., Leki, E., & Choi, S. (2022). Geographic distribution of prejudice toward African Americans: Applying the two-dimensional model. *The Journal of Social Psychology*, *162*(2), 262-279. https://doi.org/10.1080/00224545.2021.1893149

Fu, R., Gartlehner, G., Grant, M., Shamliyan, T., Sedrakyan, A., Wilt, T. J., Griffith, L., Oremus, M., Raina, P., Ismaila, A., Santaguida, P., Lau, J., & Trikalinos, T. A. (2011). Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *Journal of Clinical Epidemiology*, *64*(11), 1187-1197. https://doi.org/10.1016/j.jclinepi.2010.08.010

Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491-512. https://doi.org/10.1037/0022-3514.70.3.491

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

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2021). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2*

(updated February 2021). Cochrane. Available from
www.training.cochrane.org/handbook.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539-1558. https://doi.org/10.1002/sim.1186

*Hildebrand, L. K., Monteith, M. J., & Arriaga, X. B. (2024). The role of trust in reducing confrontation-related social costs. *Journal of Personality and Social Psychology*, *126*(2), 240-261. https://doi.org/doi.org/10.1037/pspi0000429

Hillard, A. L. (2011). *Why confronting sexism works: Applying persuasion theories to confronting sexism* [Doctoral Dissertation, University of Nebraska]. Digital Commons@University of Nebraska-Lincoln. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1035&context=psychdiss

Howell, J. L., Lofaro, N., & Ratliff, K. A. (2024). Responding to feedback about implicit bias. *Social and Personality Psychology Compass*, *18*(1). https://doi.org/10.1111/spc3.12926

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). Routledge/Taylor & Francis Group.

Hsieh, W., Faulkner, N., & Wickes, R. (2022). What reduces prejudice in the real world? A meta-analysis of prejudice reduction field experiments. *British Journal of Social Psychology*, *61*(3), 689-710. https://doi.org/10.1111/bjso.12509

Hubbard, R. R., Snipes, D. J., Perrin, P. B., Morgan, M. R., Dejesus, A., & Bhattacharyya, S. (2013). Themes in heterosexuals' responses when challenging LGBT prejudice. *Sexuality Research and Social Policy*, *10*(4), 269-278. https://doi.org/10.1007/s13178-013-0127-4

*Hyers, L. L. (1999). *Interpersonal confrontation as a means to prejudice reduction: When oppressed group members challenge the prejudices of dominant group members* [Doctoral Dissertation, The Pennsylvania State University]. The Pennsylvania State University ProQuest Dissertations. https://www.proquest.com/dissertations-theses/interpersonal-confrontation-as-means-prejudice/docview/304520282/se-2

Hyers, L. L. (2010). Alternatives to silence in face-to-face encounters with everyday heterosexism: activism on the interpersonal front. *Journal of Homosexuality*, *57*(4), 539-565. https://doi.org/10.1080/00918361003608749

Ijzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., Debruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092-1094. https://doi.org/10.1038/s41562-020-00990-w

Johnson, B. T. (2021). Toward a more transparent, rigorous, and generative psychology. *Psychological Bulletin*, *147*(1), 1-15. https://doi.org/10.1037/bul0000317

Kennedy, S. A. (1995). *Changing attitudes toward lesbian women and gay men through self-confrontation* [Doctoral Dissertation, University of North Dakota]. University of North Dakota Scholarly Commons. https://commons.und.edu/theses/3817

Klonis, S. C., Plant, E. A., & Devine, P. G. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*, *31*(9), 1237–1249. https://doi.org/10.1177/0146167205275304

*Kroeper, K. M. (2020). *Expecting prejudice confrontation to backfire: prejudice norms and misalignment between forecaster expectations and experiencer realities* [Doctoral Dissertation, Indiana University]. IUScholarWorks. https://hdl.handle.net/2022/25519

Kutlaca, M., Becker, J., & Radke, H. (2020). A hero for the outgroup, a black sheep for the ingroup: Societal perceptions of those who confront discrimination. *Journal of Experimental Social Psychology*, *88*, 103832. https://doi.org/10.1016/j.jesp.2019.103832

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York : Appleton-Century Crofts, 1970.

Lee, R. T., Perez, A. D., Boykin, C. M., & Mendoza-Denton, R. (2019). On the prevalence of racial discrimination in the United States. *PLOS ONE*, *14*(1), e0210698. https://doi.org/10.1371/journal.pone.0210698

*Lewis, T., & Yoshimura, S. M. (2017). Politeness strategies in confrontations of prejudice. *Atlantic Journal of Communication*, *25*(1), 1-16. https://doi.org/10.1080/15456870.2017.1251198

Li, A. H., Noland, E. S., & Monteith, M. J. (2024). Following prejudiced behavior, confrontation restores local anti-bias social norms. *Personality and Social Psychology Bulletin*, 1461672241229006-1461672241229006. https://doi.org/10.1177/01461672241229006

López-López, J. A., Page, M. J., Lipsey, M. W., & Higgins, J. P. T. (2018). Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research Synthesis Methods*, *9*(3), 336-351. https://doi.org/10.1002/jrsm.1310

Mahan, J. (2017, 6th June). Calling-in versus calling-out: Throwing out my activist armchair. https://www.ifwhenhow.org/calling-in-versus-calling-out-throwing-out-my-activist-armchair/

Mallett, R. K., & Wagner, D. E. (2011). The unexpectedly positive consequences of confronting sexism. *Journal of Experimental Social Psychology*, *47*(1), 215–220. https://doi.org/10.1016/j.jesp.2010.10.001

*Mark, A. Y. (2007). *Putting confrontation in its place : The influence of social norms on the effectiveness of a confrontation of prejudical bias* [Doctoral Dissertation, University of Kentucky]. Obtained from author.

Martinez, L. R., Hebl, M. R., Smith, N. A., & Sabat, I. E. (2017). Standing up and speaking out against prejudice toward gay men in the workplace. *Journal of Vocational Behavior*, *103*, 71-85. https://doi.org/10.1016/j.jvb.2017.08.001

Meyers, C., Leon, A., & Williams, A. (2020). Aggressive confrontation shapes perceptions and attitudes toward racist content online. *Group Processes & Intergroup Relations*, *23*(6), 845-862. https://doi.org/10.1177/1368430220935974

Miles, E., & Crisp, R. J. (2014). A meta-analytic test of the imagined contact hypothesis. *Group Processes & Intergroup Relations*, *17*(1), 3-26. https://doi.org/10.1177/1368430213510573

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469-485. https://doi.org/10.1037/0022-3514.65.3.469

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050. https://doi.org/10.1037/0022-3514.83.5.1029

*Monteith, M. J., Hildebrand, L. K., & Mallett, R. K. (2021). Social costs of bias confrontation are persistent but independent of bias reduction. *Manuscript in preparation*.

Monteith, M. J., Mallett, R. K., & Hildebrand, L. K. (2022). Confronting intergroup biases: Validity and impugnment as determinants of other confrontation consequences. *Advances in Experimental Social Psychology*, *66*, 1-57. https://doi.org/10.1016/bs.aesp.2022.04.001

Monteith, M. J., & Mark, A. Y. (2005). Changing one's prejudiced ways: Awareness, affect, and self-regulation. *European Review of Social Psychology*, *16*(1), 113-154. https://doi.org/10.1080/10463280500229882

Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field experiments on social media. *Current Directions in Psychological Science*, *31*(1), 69-75. https://doi.org/10.1177/09637214211054761

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. *Science (American Association for the Advancement of Science)*, *369*(6505), 866-870. https://doi.org/10.1126/science.abb3153

Munder, A. K., Becker, J. C., Ruiz, N. S., & Christ, O. (2024). At least she is doing something? Women do not prefer a woman who confronts gender-based discrimination with a distancing motive over a nonconfronter. *Psychology of Women Quarterly, 48*(2), 232-251. https://doi.org/10.1177/03616843231207052

*Munger, K. (2016). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, *39*(3), 629-649. https://doi.org/10.1007/s11109-016-9373-5

Musekiwa, A., Manda, S. O. M., Mwambi, H. G., & Chen, D.-G. (2016). Meta-analysis of effect sizes reported at multiple time points using general linear mixed model. *PLOS ONE*, *11*(10), e0164898. https://doi.org/10.1371/journal.pone.0164898

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

Noh, S., & Kaspar, V. (2003). Perceived discrimination and depression: Moderating effects of coping, acculturation, and ethnic support. *American Journal of Public Health*, *93*(2), 232-238. https://doi.org/10.2105/ajph.93.2.232

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, n71. https://doi.org/10.1136/bmj.n71

Paluck, E. L. (2011). Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, *47*(2), 350-358. https://doi.org/10.1016/j.jesp.2010.11.017

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual Review of Psychology*, *72*(1), 533-560. https://doi.org/10.1146/annurev-psych-071620-030619

*Parker, L. R., Monteith, M. J., Moss-Racusin, C. A., & Van Camp, A. R. (2018). Promoting concern about gender bias with evidence-based confrontation. *Journal of Experimental Social Psychology*, *74*, 8-23. https://doi.org/10.1016/j.jesp.2017.07.009

*Patrianakos, J. (2021). *The impact of affiliative motivation and trust on confrontations of racism* (Publication Number 3897) [Doctoral Dissertation, Loyola University Chicago]. Loyola eCommons. https://ecommons.luc.edu/luc_diss/3897

Pederson, A., Watt, S. E., & Griffiths, B. (2008). Prejudice against Australian asylum-seekers and their function: Suggestions for antiprejudice strategies [http://researchrepository.murdoch.edu.au/id/eprint/5548]. Second International Conference on "Racisms in the New World Order", University of the Sunshine Coast

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens : Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, *115*(45), 11401-11405. https://doi.org/10.1073/pnas.1721165115

Remedios, J. D., & Akhtar, M. (2019). Intersectional approaches to the study of confronting prejudice. In R. K. Mallett & M. J. Monteith (Eds.), *Confronting Prejudice and Discrimination* (pp. 179-200). Academic Press. https://doi.org/10.1016/B978-0-12-814715-3.00010-2

Rhodes, R. E., & Dickau, L. (2012). Experimental evidence for the intention-behavior relationship in the physical activity domain: A meta-analysis. *Health Psychol*, *31*(6), 724-727. https://doi.org/10.1037/a0027290

Roberts, L., Wood, C., Persson, S., Allchin, O., Simmonds-Buckley, M. (2024, June 25). Effect of Confrontation of Intergroup Bias on Reducing Intergroup Bias, and Moderators of the Effect: Systematic Review and Meta-Analysis. https://doi.org/10.17605/OSF.IO/2X3BY

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods*, *26*(2), 141-160. https://doi.org/10.1037/met0000300

Sanchez, D. T., Himmelstein, M. S., Young, D. M., Albuja, A. F., & Garcia, J. A. (2016). Confronting as autonomy promotion: Speaking up against discrimination and psychological well-being in racial minorities. *Journal of Health Psychology*, *21*(9), 1999-2007. https://doi.org/10.1177/1359105315569619

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, *470*(7335), 437-437. https://doi.org/10.1038/470437a

Sheeran, P., & Webb, T. L. (2016). The Intention-Behavior Gap. *Social and Personality Psychology Compass*, *10*(9), 503-518. https://doi.org/10.1111/spc3.12265

Shelton, J. N., Richeson, J. A., Salvatore, J., & Hill, D. M. (2006). Silence is not golden: The intrapersonal consequences of not confronting prejudice. In *Stigma and group inequality* (pp. 79-96). Psychology Press.

Shen, W., Kiger, T. B., Davies, S. E., Rasch, R. L., Simon, K. M., & Ones, D. S. (2011). Samples in applied psychology: Over a decade of tesearch in teview. *Journal of Applied Psychology*, *96*(5), 1055-1064. https://doi.org/10.1037/a0023322

*Simon, S., Mitchell, A. M., & O'Brien, L. T. (2012). *Speaking up, just to be put down: The double-edged sword of confronting racial discrimination.* Paper presented at the Society for the Psychological Study of Social Issues, Charlotte, NC.

Simon, S., & O'Brien, L. T. (2015). Confronting sexism: Exploring the effect of nonsexist credentials on the costs of target confrontations. *Sex Roles*, *73*(5-6), 245-257. https://doi.org/10.1007/s11199-015-0513-x

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*(6), 1123-1128. https://doi.org/10.1177/1745691617708630

Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus Information. *Personality & social psychology bulletin*, *27*(4), 486-496. https://doi.org/10.1177/0146167201274009

Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of public economics*, *118*, 26-40. https://doi.org/10.1016/j.jpubeco.2014.04.010

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., McAleenan, A., Reeves, B. C., Shepperd, S., Shrier, I., Stewart, L. A., Tilling, K., White, I. R., Whiting, P. F., & Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *British Medical Journal*, l4898. https://doi.org/10.1136/bmj.l4898

Stewart, K., Pedersen, A., & Paradies, Y. (2014). It's always good to help when possible, BUT...: Obstacles to bystander anti-prejudice. *The International Journal of Diversity in Education*, *13*(3), 39-53. https://doi.org/10.18848/2327-0020/CGP/v13i03/40045

Stone, J., Whitehead, J., Schmader, T., & Focella, E. (2011). Thanks for asking: Self-affirming questions reduce backlash when stigmatized targets confront prejudice. *Journal of Experimental Social Psychology*, *47*(3), 589-598. https://doi.org/https://doi.org/10.1016/j.jesp.2010.12.016

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199-214. https://doi.org/10.1037/0022-3514.68.2.199

Team, R. C. (2023). *R: A language and environment for statistical computing*. In (Version Version 4.1.2 'Bird Hippie') R Foundation for Statistical Computing. https://www.R-project.org/.

Viechtbauer, W. (2010). Conducting meta-snalyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W. (2017, October 31). *The trim and fill method has not been generalized to more complex models, so this is why this is not* [Comment on the R Special Interest Group for Meta-Analysis mailing list post "The trim and fill method for rma.mv function"]. https://stat.ethz.ch/pipermail/r-sig-meta-analysis/2017-November/000339.html

Vogels, E. A., Anderson, M., Porteus, M., Baronavski, C., Atske, S., McClain, C., Auxier, B., Perrin, A., & Ramshankar, M. (2021, 19th May 2021). *Americans and 'Cancel Culture': Where Some See Calls for Accountability, Others See Censorship, Punishment*. Pew Research Centre. Retrieved 21st April 2023 from https://www.pewresearch.org/internet/2021/05/19/americans-and-cancel-culture-where-some-see-calls-for-accountability-others-see-censorship-punishment/

Wallace, L. E., Craig, M. A., & Wegener, D. T. (2024). Biased, but expert: Trade-offs in how stigmatized versus non-stigmatized advocates are perceived and consequences for persuasion. *Journal of Experimental Social Psychology*, *110*, 104519. https://doi.org/10.1016/j.jesp.2023.104519

Watt, S. E., & Larkin, C. (2010). Prejudiced people perceive more community support for their views: The role of own, media, and peer attitudes in perceived consensus. *Journal of Applied Social Psychology*, *40*(3), 710-731. https://doi.org/10.1111/j.1559-1816.2010.00594.x

Webb, T. L., & Sheeran, P. (2006). Does changing behavioral intentions engender bahaviour change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, *132*(2), 249-268. https://doi.org/10.1037/0033-2909.132.2.249

Woods, F. A., & Ruscher, J. B. (2021). 'Calling-out' vs. 'calling-in' prejudice: Confrontation style affects inferred motive and expected outcomes. *British Journal of Social Psychology*, *60*(1), 50-73. https://doi.org/10.1111/bjso.12405

Woodzicka, J. A., Mallett, R. K., Hendricks, S., & Pruitt, A. V. (2015). It's just a (sexist) joke: Comparing reactions to sexist versus racist communications. *Humor*, *28*(2), 289-309. https://doi.org/10.1515/humor-2015-0025

Woodzicka, J. A., Mallett, R. K., & Melchiori, K. J. (2020). Gender differences in using humor to respond to sexist jokes. *Humor*, *33*(2), 219–238. https://doi.org/10.1515/humor-2019-0018