2022

Not Such a Complainer Anymore: Confrontation That Signals a Growth Mindset Can Attenuate Backlash

Aneeta Rattan
London Business School

Kathryn Kroeper
Sacred Heart University, kroeperk@sacredheart.edu

Rachel Arnett
Wharton School, University of Pennsylvania

Xanni Brown
Yale University

Mary Murphy
Indiana University

Recommended Citation

Not Such a Complainer Anymore:

Confrontation that Signals a Growth Mindset can Attenuate Backlash

Date Resubmitted: March 28, 2022

Abstract (250 words)

We report the first investigation of whether observers draw information about mindsets from behaviour, specifically prejudice confrontation. We tested two questions across 10 studies ($N = 3168$). First, would people who observe someone confront a biased comment (vs. remain silent) see them as endorsing more growth (vs. fixed) mindsets about prejudice and bias? If so, would the growth mindset perceptions that arise from confrontation (vs. remaining silent) attenuate the backlash that observers exhibit against confronters? We investigated these questions using scenarios (Studies 1, 2a-b, 4, 5a-d), naturalistic confrontations of national, race, and gender stereotypes reported retrospectively (Study 3), and an in-person laboratory experiment of actual confrontations of racial bias (Study 6). Correlational and experimental methods yielded support for our core hypotheses: People spontaneously imbue someone who confronts a biased comment with more growth mindset beliefs about prejudice and bias (Studies 1, 2a-b, 4, 6), regardless of whether participants observe the confrontation (Studies 1, 2a-b, 5a-d) or are being confronted themselves (Studies 2a-4, 6). The growth mindset perceptions arising from these confrontations suppress backlash, assessed by classic interpersonal perceptions (Studies 4-5) and judgments of interpersonal warmth and willingness to interact again in the future (Study 6), both when the confronter was a target of the biased behaviour (Studies 1-5), and when they were an ally (Study 6), in both correlational studies (Study 3-4) and when growth mindset (about personality, Study 5; about prejudice, Study 6) was manipulated, confirming causality. We discuss implications for the study of mindsets, confrontation, and intergroup relations.

Keywords: backlash; confrontation; lay theories; mindset; prejudice

Introduction

Imagine a woman confronting her male work colleague for saying that women are simply too emotional to become great leaders within the organization; likewise, imagine a Black person speaking out to disagree with their White co-worker's comment that they must have been a 'diversity hire' at the company, or imagine a majority group member ally confronting in either situation. Across these scenarios, targets of bias and allies are positioned to confront biased comments targeting social identities—and, in doing so, they risk inflaming social and professional backlash. Indeed, longstanding and well-replicated phenomena in bias confrontation scholarship show that racial minorities and women both anticipate and receive harsh, negative social evaluations for speaking out (i.e., backlash, Kaiser & Miller, 2001; Schultz & Maddox, 2013). While majority group member allies may receive fewer negative evaluations than racially minoritized populations and women, they too still worry about and experience backlash (Gervais & Hillard, 2014; Kutlaca et al., 2020). Troublingly, anticipation of these social and professional costs reduces people's likelihood of speaking out (Ashburn-Nardo et al., 2008, 2014; Shelton & Stewart, 2004), meaning bias expression is likely to continue. Surprisingly little research has found ways to mitigate the backlash that people exhibit toward those who confront biased comments (but see Stone et al., 2011).

In the current research, we sought to investigate, first, whether in addition to the negative interpersonal perceptions that arise when racially minoritized people or women confront biased comments, there might be an as-yet-undocumented *positive* interpersonal perception that also arises. Specifically, we hypothesized that confrontation might signal that the person who speaks up to address a biased comment holds more of a growth, rather than fixed, mindset about prejudice and bias. If so, we proposed to test whether these growth mindset perceptions could function psychologically to suppress interpersonal backlash effects. That is, we examined whether the growth mindset attributions afforded to confronters

might mitigate backlash against racially minoritized people and women when they speak out on behalf of their group, as well as whether this protection extends to majority group member allies when they speak out.

Can Confrontation Cue Perceptions of a Growth Mindset?

Mindsets, or lay theories, describe people's beliefs about whether human characteristics (e.g., intelligence, personality, prejudice) are fixed or malleable (Dweck, 1999). In the context of bias and prejudice, an individual who views prejudice as a malleable trait that can change with intention and effort would be considered to hold a relatively growth mindset; by contrast, an individual who views prejudice to be a fixed characteristic, that does not meaningfully change, would be considered to hold a relatively fixed mindset (Carr et al., 2012; Neel & Shapiro, 2012). Decades of research establishing the study of lay theories about malleability has focused on how an individual's fixed versus growth mindset shapes their goals, emotions, persistence, and performance (for reviews, see Dweck & Yeager, 2019; Rattan & Ozgumus, 2021).

However, more recent research has begun to explore people's perceptions of others' mindsets (or their metacognitions about others' mindsets). For example, research has explored perceptions of the generalized mindsets imbued in classrooms (Good et al., 2012) and institutions (e.g., organizational mindsets, Canning et al., 2020; Murphy & Reeves, 2019), showing that these mindset perceptions shape students' sense of belonging and employees' perceptions of the company culture, respectively. Other studies have, to date, explored how students perceive their teacher's mindsets about student intelligence, finding that these perceived teacher mindsets impact students' motivation and performance (Canning et al., 2021; Fuesting et al., 2019; LaCosse et al., 2021; Muenks et al., 2020; Rattan et al., 2018). Research on mindset perceptions has relatedly found that people hold mindsets of certain social groups (e.g., liberals versus conservatives, younger versus older people) as

relatively more malleable than fixed, respectively (Lassetter & Neel, 2019; Neel & Lassetter, 2015). In sum, this burgeoning direction of mindset research shows that, in addition to holding their own mindset beliefs, people also mentalize about the mindsets of individuals, groups, and institutions. Moreover, this work reveals that people's mindset perceptions shape self-relevant outcomes such as feelings and behaviour.

No research, to our knowledge, has explored whether behaviours—like confronting a biased comment—might shape mindset perceptions. In the current research, we extend the study of mindset perceptions, and advance the core theory of mindsets, in three ways. Our work is the first, to our knowledge, to test whether a specific behavioural trigger (i.e., confrontation) creates perceptions of another's mindset beliefs. Second, we advance the study of mindsets in two nascent areas of mindset research—extending work on mindsets about prejudice and bias (Carr et al., 2012; Neel & Shapiro, 2012; also see Hennes et al., 2018; Simon et al., 2019) as well as the study of mindsets and confrontation (Rattan & Dweck, 2010, 2018). Third, we shift our focus away from studying how perceiving others' mindsets shapes self-relevant outcomes (e.g., one's own feelings and behaviours). Instead, our work extends this field toward studying how people's perceptions of other's mindsets shape their treatment of those individuals. Specifically, we investigate the hypothesis that the confrontation of biased behaviour can cue perceptions that an individual endorses more growth (vs. fixed) mindset beliefs about prejudice and bias, which, in turn, suppresses perceivers' likelihood of exhibiting backlash in their interpersonal perceptions or behaviours.

Why would we expect confrontation to cue growth mindset perceptions? Past research has documented that racially minoritized people and women who endorse more growth (vs. fixed) mindset beliefs are more likely to anticipate confronting and actually confront a biased comment (Rattan & Dweck, 2010). Likewise, people who endorse more growth (vs. fixed) mindset beliefs are more likely to engage in potentially uncomfortable discussions about

racial issues (Carr et al., 2012; Neel & Shapiro, 2012). This suggests that people who hold growth mindsets are often motivated to engage with someone else's bias, or issues of bias more generally. Considering these findings alongside a separate body of social cognition research showing that people are surprisingly good at spontaneously reading others' intentions from their actions (Becchio, 2012; Slepian et al, 2013; Uleman et al., 1996; also see Woods & Ruscher, 2021), we hypothesized that confrontation may indicate to perceivers that a person who speaks up holds a growth, rather than fixed, mindset about prejudice and bias. In other words, when someone speaks up to confront a biased statement, observers imbue their action with an assumed intention of changing the biased actor's beliefs and/or their behaviour. Speaking out to disagree with biased comments, therefore, may suggest to observers that the confronter believes prejudice and biases can change (otherwise, they likely wouldn't bother speaking up); conversely, the act of remaining silent, despite disagreeing with a biased comment, may suggest that the *non*-confronter believes that prejudice and bias cannot change.

Of course, counter to the theory described above, two counterfactuals exist which make it especially important to empirically test what (if anything) people infer about the mindset beliefs of confronters from observing their behaviour. First, it could be that confrontation behaviour does not cue mindsets at all. Instead, people might read confrontation as simply an expression of how offensive the confronter found the comment to be—confronting only when they find the bias to be particularly offensive—and, in such a case, we should find no differences in people's perceptions of confronters' and non-confronters' mindsets. Second, it is alternatively possible that a confrontation could signal *fixed* mindset beliefs—insofar as it is viewed as labelling or shaming 'bad,' 'irredeemable' actors. Acknowledging these alternative possibilities highlights the importance of investigating whether, and how, the act of confrontation may shape mindset perceptions.

Can Confrontation that Cues a Growth Mindset Suppress Backlash?

If confrontation signals a growth mindset, as we theorize, we wondered whether this perception could undercut (i.e., suppress) the well-known negative interpersonal consequences that follow when people speak up to challenge bias (Czopp & Monteith, 2003; Kaiser & Miller, 2001). While research has found that confronting a biased comment can effectively reduce the reoccurrence of bias (Czopp et al., 2006; Drury & Kaiser, 2014; Focella et al., 2015; Gulker et al., 2013) and establish anti-biased norms in the context (Blanchard et al., 1994; Kroeper, 2020; Paluck, 2011), studies have also shown that these benefits arise at the expense of those who speak out. That is, outside observers and confronted individuals alike dislike racially minoritized individuals and women who speak out, even to overtly biased comments (Rasinski & Czopp, 2010). This phenomenon has been called 'backlash,' whereby people judge racially minoritized people and women who confront (vs. remain silent) as hypersensitive complainers, view them less favourably, and avoid or withdraw from future interactions with them (Czopp & Monteith, 2003; Gulker et al., 2013; Kaiser & Miller, 2001). Notably, even though majority group members rarely confront (Kawakami et al., 2009; Karmali et al., 2017), when they do speak out, they too experience backlash, though it is generally less severe than what is experienced by racially minoritized individuals and women (Cadieaux and Chasteen, 2015; Gervais & Hillard, 2014; Gulker et al., 2013). It is notable that studies investigating interpersonal perceptions of confronters have so consistently documented the largely negative effects of speaking out, in particular backlash.

In this context, our investigation advances the study of prejudice confrontation in a new direction—proposing that in addition to the well-documented negative interpersonal evaluation of backlash, a seemingly positive perception, that the confronter has a more growth mindset, may also emerge. We propose and test whether these growth mindset

perceptions may have the power, psychologically, to suppress backlash effects. All else being equal, previous research suggests that confronting bias (vs. remaining silent) will increase the interpersonal backlash perceivers direct toward an actor (e.g., Czopp & Monteith, 2003). But, as we are the first to argue, confronting bias might also increase the perception that an actor endorses growth mindset beliefs about bias and prejudice. Because perceiving confronters as holding a growth mindset presumes a positive intention behind their confrontation, and given the large body of research showing that intentionality shapes how people respond to even negative behaviour (Adams, 2016), we theorized that to the extent perceivers imbue confronters with growth mindset beliefs, the level of backlash they exhibit should be suppressed (cf. Carlston & Skowronski, 1994; Carlston et al., 1995; Uleman et al., 1992, 1996). In other words, in line with statistical suppression (Hayes, 2018; MacKinnon et al., 2000), we expect confrontation to elicit weakened backlash in the presence of growth mindset perceptions.

While the early part of our investigation tests only people's perceptions of confronters' mindset beliefs, the latter part of our investigation tested this suppression possibility by, first, measuring and, second, directly manipulating people's perceptions of confronters' mindset beliefs and examining whether the presumption of growth mindset beliefs suppresses the degree of backlash people exhibit. Across studies, we explore two types of backlash—the field-standard interpersonal perceptions of backlash as seeing a confronter as more of a complainer and less likeable (Kaiser & Miller, 2001), as well as the social consequences of such negativity, indexed by lowered warmth perceptions (Hennessey, 2018), and less of a desire to interact with the confronter in the future (Kowalski & Erickson, 1997).

Actor vs. Observer Differences in Mindset Perceptions?

To summarize, we sought to investigate two questions. First, would people who see someone confront a biased comment (vs. remain silent) perceive the confronter as endorsing more growth (vs. fixed) mindset beliefs about prejudice and bias? If so, would the growth mindset perceptions that arise from confrontation (vs. remaining silent) suppress the level of backlash that people exhibit? Of course, there are two ways to experience a confrontation of a biased comment. One can either be a third-party observer, watching and evaluating a person subjected to *someone else's* biased comment. Or, one can be an actor in the interaction themselves, watching and evaluating a person responding to *one's own* biased comment. While studying our focal research questions, we tested for differences between these actor-observer perspectives (Jones & Nisbett, 1971; Malle, 2006). Our goals were twofold—first we hoped to provide a full investigation of the processes we study, and second, we hoped to test whether there might be limits to people's willingness to afford someone a growth mindset, in line with the self-serving patterns documented in past actor-observer research (Malle, 2006). Given the discomfort and irritation people exhibit when they are confronted for a biased comment (Czopp et al., 2006), we speculated in our early, exploratory approach that the actor (vs. observer) perspective would arouse a self-threat that would motivate individuals to denigrate their confronter and, therefore, be reluctant to imbue positive attributes upon them, including a growth mindset. That is, in our initial investigation, we thought it was possible that people's self-defensiveness at the thought of being confronted for their own biased comment might prevent them from affording the person who speaks up a growth mindset. In these early studies, we predicted that confrontation would cue a growth mindset only for those in the observer perspective and not for those in the actor perspective. As we report below, we did not find support for these predictions—we found no evidence of actor/observer differences in growth mindset ascriptions. This suggested that we initially

underestimated the ubiquity of these mindset ascriptions in the confrontation context and,

therefore, in our later studies we stopped predicting differences along this dimension.

Overview of Studies

This manuscript reports the results of 10 studies (total $N = 3168$) which sought to

explore the dynamics of confrontation, mindset perceptions, and backlash. Studies 1, 2a, and

2b are experimental investigations of our first hypothesis: that people would perceive a

woman who confronts (vs. stays silent) in response to a sexist comment as endorsing more

growth mindset beliefs about prejudice and bias. Study 1 places participants in an observer

role, while Studies 2a and 2b experimentally vary whether participants were in the actor or

observer role. Study 3 explores these metacognitions about others' mindsets in the context of

naturalistic confrontations, where participants recounted their own experiences of being

confronted for actual biased comments they made about someone's nationality, race, gender,

or other identity. Study 3 uses correlational methods to test our second hypothesis: that

imbuing a confronter with growth mindset beliefs suppresses the expression of backlash.

Study 4 conceptually replicates these effects but tests for causality by again experimentally

manipulating confrontation. Study 4 also begins to explore the potential consequences of

growth mindset perceptions and backlash suppression, specifically on participants'

willingness to interact with the confronter again in the future. Study 5 presents a mini-meta-

analysis of Studies 5a-5d, all of which use the same 2 (confrontation condition: confrontation

vs. no confrontation) × 2 (mindset condition: fixed versus growth) experimental design to

test whether a confronter explicitly endorsing a growth mindset suppresses the degree of

backlash people exhibit. Study 6 turns to a stronger situation, in which participants in the

course of a lab study are actually confronted for racial bias they express by their purported

partner, a White ally. We experimentally manipulate whether the confrontation message itself

communicates a growth mindset, a fixed mindset, or no mindset (control condition) to

evaluate whether the growth mindset suppresses backlash and thus improves participants' willingness to interact with the confronter again. Thus, across the studies in this manuscript, we test the hypotheses, first, that confronting bias can cue people to perceive that the confronter endorses more growth mindset beliefs about prejudice and bias, and second that growth mindset perceptions in a confrontation in turn attenuate backlash against the confronter.

Notably, racism is often seen as more offensive and confrontation-worthy than sexism (Parker, et al., 2018; Woodizcka et al., 2015), which makes it all the more important that we test the psychological processes under investigation across bias against multiple stigmatized social identities to examine the generalizability of these effects. Across the studies, in an effort to be thorough in examining these hypotheses, we use both correlational and experimental methods, we investigate both scenarios and real-life confrontations, we test these dynamics across both actor and observer perspectives, considering evaluations of both targets of bias and allies who confront, while experimentally manipulating both confrontation and mindset in the contexts of both sexism and racism in order to provide robust, convergent, causal evidence for any conclusions.

Data for each study was always collected in a single wave and were analysed after the study closed. All conditions, materials, and methods are reported in full. Participants in chronologically earlier studies were excluded from participating in later studies. We note that all studies prior to Study 6 had internal lab documentation to establish a priori predictions but were not pre-registered. Study 6 was pre-registered (anonymized url: https://osf.io/j7xbz/?view_only=594a6feb70924d3393583240fe44eb29).

Study 1

Study 1 was an exploratory investigation of whether prejudice confrontation might signal to observers that the confronter holds growth mindset beliefs. We tested this using a 2 condition (confront vs. no confront) between-subjects design.

Method

Participants

We aimed to achieve a minimum of $N = 50$ per cell, even with a priori exclusions, so we posted a hit for 125 people on Amazon Mechanical Turk (MTurk). Sixty-eight men, 56 women, and 1 agender adult participated ($M_{Age} = 32.62$, $SD = 10.24$); these participants self-identified race/ethnicity as: 4 African American, 102 European American/White, 4 Latinx American, 10 Asian American, 1 African American-European American/White-Native American, 2 Latinx American-European American/White, 2 Asian American-European American. Twenty-five participants were excluded from all analyses for failing attention and memory check questions (12 from the no confrontation condition and 13 from the confrontation condition), resulting in a final sample size of 100. A sensitivity power analysis for a two-tailed, independent samples t-test conducted post-hoc using G*power indicates $N = 100$ would detect an effect size $d = .57$ with 80% power and alpha = .05.

Procedure

After providing informed consent, participants read a scenario about an incident between co-workers involving a gender biased statement (adapted from Rattan & Dweck, 2010, 2018):

> One day recently, Emily took a break to get a coffee and found herself speaking with John. John is a new employee whose name Emily has heard a few times but who she had not met before. John is at the same level in the organization as Emily. Although he works in a different area from Emily, his group has about the same status within the organization. <page break>

Emily and John began discussing his first impressions of the company. After mentioning a few aspects of the position that he likes, John said "You know, I'm really surprised at the types of people who are working here. When you get to the top level—a company like this—you expect only the best people. I mean, I think they must be hiring associates just for diversity reasons. With all the women here, I wonder how long this company will stay on top."

Participants were then randomly assigned to the prejudice confrontation condition. In the *no confrontation condition*, they read, "After a moment of silence, Emily did not respond to John about his statement. Although Emily disagreed with his comment, she said nothing." And in the *confrontation condition*, they read, "After a moment of silence, Emily confronted John about his statement. Emily clearly expressed her disagreement with his comment in a calm but firm voice." Next, participants completed the following measures.

Perceived Mindset of the Confronter. In order to assess whether participants saw confrontation as indicative of growth mindset beliefs about prejudice, they were asked to complete an adapted version of the 6-item Theories of Prejudice Scale, which measures people's beliefs about the malleability of prejudice (from Carr et al., 2012). Scale items were revised to refer to the mindset beliefs of the woman in the scenario (rather than one's own personal mindset beliefs, as in the original scale). That is, participants indicated their agreement with statements such as, "Emily believes that no matter who somebody is, they can always become a lot less prejudiced" and "Emily believes that people have a certain amount of prejudice and they can't really change that" (reverse-scored), on a 6-point scale from "*very strongly disagree*" to "*very strongly agree*" with higher numbers indicating that Emily was perceived to endorse more growth mindset beliefs about prejudice ($\alpha = .91$).

Confronter's Openness to Future Interactions with Perpetrator. Participants rated how much they thought, "Emily would be willing to interact with John again in the future."

and, "Emily would avoid future interactions with John at all costs." (reverse-scored) on a 7-point scale from "*strongly disagree*" to "*strongly agree*" ($r = .66$, $p < .001$).

Attention and Manipulation Checks. Participants answered 2 multiple choice attention check questions, which tested whether they remembered information from the scenario they had read, and 1 manipulation check item which asked them to select an accurate statement (out of 4) about Emily's response (whether she disagreed and spoke up, disagreed and stayed silent, agreed and spoke up, agreed and stayed silent).

In addition, we measured how offensive participants thought Emily had found the comment, how offensive participants found the comment, how much they thought Emily's behaviour was effective at changing John's beliefs, and participants' personal mindset beliefs about prejudice. For the sake of brevity, and given that the effects reported below hold controlling for each of these variables in turn, we report the details and results of these measures in the SOM. Finally, participants completed a standard demographics form, were debriefed, and were paid.

Results

We conducted independent samples t-tests of confrontation condition (confrontation present vs. absent) on each of the dependent variables.

Perceived Mindset of the Confronter. In support of the hypothesis that prejudice confrontation may signal growth mindset beliefs on the part of the confronter, participants in the confrontation condition rated Emily as endorsing a significantly more growth mindset about prejudice ($M = 4.36$, $SD = .89$) than did participants in the no confrontation condition ($M = 3.28$, $SD = .92$), $t(98) = -5.91$, $p < .001$, Cohen's $d = 1.19$ (see Figure 1 panel a).

Confronter's Openness to Future Interactions with Perpetrator. The difference in perceptions of Emily's openness to future interactions with John as a function of condition

was not significant, $t(85.68) = -0.17$, $p = .10$, Cohen's $d = .34$, confrontation condition ($M = 2.50$, $SD = 0.96$), no confrontation condition ($M = 2.86$, $SD = 1.21$).

Discussion

Study 1 offers initial evidence that people can perceive confronters of gender bias to endorse more growth mindset beliefs about prejudice. When the confronter spoke up (vs. did not) in response to the biased statement, participants perceived her to believe that people can change their level of prejudice more. Further, as we report in detail in the SOM, this effect holds even after controlling for how offensive participants thought the woman would find the gender biased statement, ruling out the possibility that perceptions of the comments' offensiveness are what drive mindset perceptions. We did not find differences by confrontation condition on Emily's (the confronter's) openness to future interactions with John (the person who made the biased comment). Because this exploratory measure did not yield conclusive results, we did not pursue it further. Instead, in the later studies we return to the construct of willingness to engage in future interactions, but in those studies, we examine this in the context of backlash, or observers' willingness to interact with the confronter again (i.e., the opposite of what was measured in this study).

Figure 1

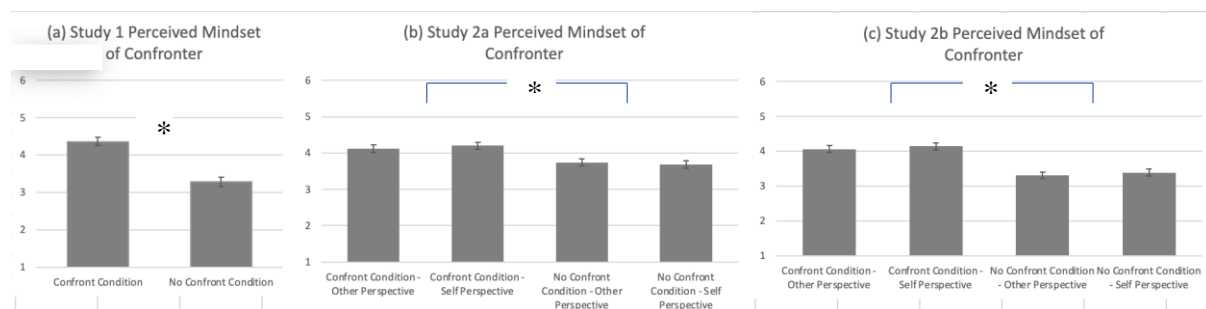*Figure 1. Summary of means (and standard errors) for perceptions of the confronter's mindset by condition across Studies 1 (Panel a), 2a (Panel b), and 2b (Panel c). The only reliable main effect to emerge as significant is the main effect of confrontation condition – participants randomly assigned to read that someone confronts (vs. remains silent) rate her as holding a more growth (vs. fixed) mindset about prejudice and bias.*

Study 2a & 2b

Study 2a-b sought to replicate and extend Study 1 in multiple ways. First, we tested whether, in addition to third-party observers (as shown in Study 1), people who imagine *themselves* being confronted (i.e., as being *actors* in the scenario) will likewise perceive confrontation to signal growth mindset beliefs on the part of the confronter. Second, this study used a different overtly gender biased statement in order to ensure that any effects found are not incidental to the statement used in Study 1. Third, the scenario offered more specificity to make it clear exactly which comment Emily disagreed with. Because the only real difference between Studies 2a and 2b is the sample (from Prolific Academic and MTurk, respectively), we present them together.

Studies 2a and 2b employ a 2 (Perspective Condition: actor vs. observer) × 2 (Confrontation Condition: no confrontation vs. confrontation) between-subjects design. Our a priori hypothesis was that the confrontation (vs. no confrontation) condition would again yield more growth mindset perceptions, but only in the context of someone else being confronted (observer condition) and not the self being confronted (actor condition). Given that previous research suggests that people often do not like individuals who confront them (Parker et al., 2018; Schultz & Maddox, 2013), we initially expected people's self-defensiveness at the thought of being confronted for a biased comment to prevent them from affording the person who speaks up more growth mindset beliefs. Indeed, this pattern of results would have been an interesting limitation on the findings observed in Study 1. As noted above, this a priori hypothesis was not supported by the data and thus in the studies that follow we no longer predict actor-observer perspective differences.

Method

Participants

16

For each sample, we aimed to achieve a minimum of $N = 75$ per cell, even with a priori exclusions, so we posted two studies: one requesting 400 people on Prolific Academic for Study 2a and another study requesting 400 people on MTurk (we received 402 responses) for Study 2b.

Study 2a. To ensure that the two perspective conditions were equally realistic (being confronted for biased behaviour that was committed by the self vs. by another person), we offered the study only to men using pre-qualification criteria on Prolific. Participants were 396 men, 3 women, and 1 demimale, $M_{Age} = 32.62$, $SD = 10.24$, who self-identified their race/ethnicity as: 22 African American, 291 European American/White, 17 Latinx American, 35 Asian American, 4 Native American, 3 African American-European American/White, 3 African American-Latinx American, 1 African American-Asian American, 8 European American/White-Latinx American, 3 European American/White-Native American, 6 European American/White-Asian American, 1 European American/White-Middle Eastern American, 1 European American/White-Latinx American-Native American-Middle Eastern American, 1 biracial, 3 mixed, 1 American. Seventy-two participants were excluded prior to data analysis for either not meeting participation criteria (i.e., not self-identifying as men on the survey) or failing the manipulation/attention checks, leaving 328 participants in the final sample (Actor, No confrontation condition $N = 80$; Actor, Confrontation condition $N = 86$; Observer, No confrontation condition $N = 83$; Observer, Confrontation condition $N = 79$). A sensitivity power analysis for an ANOVA (fixed effects, special, main effects and interactions) conducted post-hoc using G*power indicates that $N = 328$ would detect an effect size $f = .16$ with 80% power and alpha = .05.

Study 2b. We again only intended to recruit men, but an oversight meant that the study was left open to men and women. Participants were 199 men, 201 women, 1 FTM, and 1 nonbinary, $M_{Age} = 37.46$, $SD = 12.25$, who self-identified their race/ethnicity as: 23 African

American, 303 European American/White, 21 Latinx American, 33 Asian American, 3

Native American, 1 European American/White-African American, 1 African American-

Native American, 1 Latinx American-Native American, 1 European American/White-Native

American, 2 European American/White-Middle Eastern American, 1 Asian American-

Middle Eastern American, 2 African American-Native American-Latinx American, 1

European American-African American-Asian American, 1 self-described "human." Sixty-

four participants were excluded prior to data analysis for failing the manipulation/attention

checks, leaving 338 participants in the final sample (Actor, No confrontation condition $N =$

78; Actor, Confrontation condition $N = 85$; Observer, No confrontation condition $N = 92$;

Observer, Confrontation condition $N = 83$). A sensitivity power analysis for an ANOVA

(fixed effects, special, main effects and interactions) conducted post-hoc using G*power

indicates that $N = 338$ would detect an effect size $f = .15$ with 80% power and alpha = .05.

Procedure for Studies 2a and 2b

After providing informed consent, participants in both Study 2a and 2b were

randomly assigned to the actor or observer condition. Those randomly assigned to the actor

condition read a scenario that placed themselves in the story, while those randomly assigned

to the observer condition read a scenario that referred to John (as in Study 1). Participants

read:

>Emily works within a large professional services organization. She loves her job and
>is widely considered a high performer. Her team is responsible for the company's presence in
>the media and deals with journalists and social media on a daily basis. <break>
>
>One day recently, Emily took a break to get a coffee and found herself speaking with
>*you (John)*. *You (John)* are an employee whose name Emily has heard a few times but who
>she had not met before. *You (John)* are at the same level in the organization as Emily.
>Although *you (he)* work in a different area from Emily, *your (his)* group has about the same
>status within the organization. Emily and *you (John)* begin discussing *your (John's)*

impressions of the company. After mentioning a few aspects of the position that *you (he)* like*(s)*, *you (he)* say*(s)* "I am just so glad I didn't end up on a team with a woman manager. Women are just too emotional to manage teams effectively, and those teams will just never rise to the top or be stars."

Participants were then randomly assigned to the prejudice confrontation condition. In the *no confrontation condition*, they read, "After a moment of silence, Emily does not respond to this statement. Although she privately disagrees with the comment about women's ability to manage teams effectively and star performance, she gets her coffee and says she needs to head back to her desk." And in the *confrontation condition*, they read, "After a moment of silence, Emily confronts this statement. Emily clearly expresses her disagreement with the comment about women's ability to manage teams effectively and star performance in a calm but firm voice." Next, participants completed the following measures.

Perceived Mindset of the Confronter. Participants' ratings of Emily's mindset beliefs about prejudice were measured as described in Study 1 (Study 2a, $\alpha = .88$; Study 2b, $\alpha = .90$).

Attention and Manipulation Checks. Participants answered 2 multiple choice attention check questions, which tested whether they remembered information from the scenario they had read. They also answered 1 manipulation check item which asked them to accurately identify if they were in the actor or observer condition, and 1 which asked them to select an accurate statement (out of 4) about Emily's response. Participants who answered any of these questions incorrectly were excluded prior to data analysis, which represented a priori exclusions.

In addition, we measured how offensive participants thought Emily had found the comment in Studies 2a and 2b, how offensive participants personally found the comment in Study 2a, how likely participants thought they were to make the comment in the actor

condition of Study 2a, and participants' agreement with the comment across conditions in Study 2b. For the sake of brevity, and given that the results are consistent controlling for each of these measures in turn, we report the details and results of these measures in the SOM. Finally, participants completed a standard demographics form, were debriefed, and paid. Also see SOM for analyses moderated by gender for Study 2b, which also do not change the interpretation of the results reported below.

Results

We conducted 2 (Perspective Condition: actor vs. observer) × 2 (Confrontation Condition: confront vs. no confrontation) ANOVAs on the dependent variables.

Study 2a Perceived Mindset of the Confronter. No main effect of perspective emerged, $F(1, 327) = 0.02$, $p = .90$, $n_p^2 < .001$. Like Study 1, participants in the confrontation condition ($M = 4.16$, $SD = 0.83$) perceived Emily to endorse more growth mindset beliefs about prejudice than participants in the no confrontation condition ($M = 3.71$, $SD = 0.96$), $F(1, 327) = 20.67$, $p < .001$, $n_p^2 = .06$. Contrary to the hypothesis, there was no perspective condition × confrontation condition interaction on participants' ratings of Emily's mindset beliefs, $F(1, 327) = 0.53$, $p = .47$, $n_p^2 = .002$ (see Figure 1 panel b).

Study 2b Perceived Mindset of the Confronter. The results were consistent with the results of Study 2a. No main effect of perspective emerged, $F(1, 337) = 0.59$, $p = .44$, $n_p^2 = .002$. Again, participants in the confrontation condition ($M = 4.10$, $SD = 0.90$) perceived Emily to endorse more growth mindset beliefs about prejudice than participants in the no confrontation condition ($M = 3.35$, $SD = 0.88$), $F(1, 337) = 57.74$, $p < .001$, $n_p^2 = .15$. Contrary to the hypothesis, there was no perspective condition × confrontation condition interaction on participants' ratings of Emily's mindset beliefs, $F(1, 337) = 0.002$, $p = .96$, $n_p^2 < .001$ (see Figure 1 panel c).

Discussion

Contrary to our predictions, we did not find an interaction of actor-observer perspective and confrontation in Studies 2a and 2b. Instead, a consistent and reliable pattern emerged: participants who read that either they or another person had made a gender biased statement imbued Emily with more growth mindset beliefs about prejudice when she confronted, compared to when she remained silent. Taken together, this pattern now replicates across two types of gender biased comments: biased comments about hiring for diversity purposes (in Study 1) and biased comments about women's emotionality (in Study 2 a-b). Because we found no evidence of actor-observer differences in the growth mindset perceptions that follow from confrontation, we no longer hypothesize about or test these differences in the studies that follow. Further, as we report in detail in the SOM, participants did not vary in their perceived likelihood of making the biased statement (Study 2a) or agreement with it (Study 2b), suggesting that participants in the actor (vs. observer) condition did not reject the scenario.

Study 3

Why should we care that confrontation shapes people's perceptions of the confronter's mindset beliefs? What are the consequences of viewing confrontation as stemming from a belief in people's ability to change and grow? Study 3 begins to explore these questions by eliciting retrospective accounts of actual bias confrontation and moving from an online to an in-person sample. In this study, students recalled a situation in which they expressed a stereotype about someone's nationality, race, gender, or other personal identity and had been confronted for it. They then reported their perceptions of the mindset beliefs of their confronter and completed a classic scale of interpersonal backlash.

Method

Participants

21

The study was embedded in a survey conducted for teaching purposes. Because the main content of the survey would be taught in class, all students in the first year of an MBA program completed the measures. However, only students who consented for their responses to be used in academic research ($N = 406$; 96%) were retained for the analyses reported here. Participants were 150 women, 223 men, and 2 nonbinary individuals. No other demographic characteristics were assessed. A sensitivity power analysis for a bivariate correlation conducted post-hoc using G*power indicates that $N = 159$ would detect an effect size $r = .22$, and $N = 74$ would detect an effect size $r = .32$ with 80% power and alpha = .05.

Procedure

Participants were enrolled in an international MBA program, in which 63 nationalities are represented; no single national context represents more than 20% of the student population. Given the national diversity, and that engagement with this type of diversity is a frequent topic of conversation, we asked students to recall a time when they were confronted for expressing a stereotype. First, they read, "Everyone holds stereotypes and assumptions about other social groups (e.g., groups based on nationality, race, gender, or sexual orientation, etc.). As a diverse environment, [school name] provides the opportunity for people to speak up and address each other's stereotypes or assumptions about other social groups. Please think back over the interactions you have had at [school name] thus far. Can you think of a specific interaction or conversation in which someone spoke up to address a stereotype or assumption that you expressed about another social group?"

Participants who selected "yes" moved on to complete the following measures ($N = 159$), while participants who selected "no" skipped to other measures for teaching purposes ($N = 247$).

Free response. In order to confirm that they were thinking of an appropriate situation, participants were asked to describe the stereotype they expressed and what was said to them

in response. They were also asked to enter the initials of the person who confronted them in a textbox, which were piped into the following measures.

Perceived Mindset of the Confronter. Participants rated their perceptions of their confronter's mindset beliefs using the measure described in Study 1; however, in this study, "Emily" was replaced with the initials of the person who confronted them (α = .89; e.g., "A.R. believes that no matter who somebody is, they can always become a lot less prejudiced.").

Backlash. Participants then completed the classic, field-standard measure of backlash against confronters from Kaiser & Miller (2001), which is composed of two subscales known as complainer and favourability. Participants rated the extent to which they perceived their confronter to be hypersensitive, irritating, a complainer, a troublemaker, emotional, and argumentative, which were averaged to form the complainer subscale (α = .85), and likeable, friendly, honest, easy to get along with, intelligent, independent, responsible, optimistic, respectable, considerate, nice to converse with, made a good impression, would be a good friend, would be a good coworker, has a good personality, and has a strong work ethic, which were averaged to form the favourability subscale (α = .97), on a 7-point scale from "*not at all*" to "*very much*." Backlash is characterized by rating an individual as simultaneously both more of a complainer and less favourably (Kaiser & Miller, 2001).

<div align="center">

Results

</div>

Unexpectedly, many participants did not complete the free response measure, though they completed all other measures for this study. We attribute this to survey fatigue, given that this survey was placed last, after a 15-minute teaching-related survey. Given this, we present the results among all who chose "yes" when asked to recall a time when they were confronted for expressing a stereotype (N = 159, 38.2% of respondents) and separately among those who we could confirm shared a qualifying situation, as determined by a coder

who coded for whether participants described an incident in which they had communicated bias and been confronted ($N = 74$). As would be expected in this research setting, the coding showed that most of the biased comments communicated stereotypes about nationality, which overlapped with (and in some cases are not separable from) racial stereotypes, followed by gender stereotypes.

First, we examined the results among those who indicated they could recall a time when they had been confronted for expressing a stereotype ($N = 159$, see Table 1 for means, standard deviations, and correlations). This correlational analysis revealed that the more participants imbued their confronter with growth mindset beliefs about prejudice, the less they demonstrated backlash on both subscales of the measure—rating the confronter as less of a complainer, $r = -.24$, $p = .002$, and more favourably, $r = .34$, $p < .001$.

Next, we examined the results among those whose clearly expressed a stereotype and had been confronted (as coded by the coder; $N = 74$, see Table 1 for means, standard deviations, and correlations). Among these participants, perceptions of their confronter's growth mindset beliefs about prejudice only predicted backlash on one subscale. That is, these perceived mindset beliefs were unrelated to participants' perceptions of the confronter on the complainer subscale of the backlash measure, $r = -.12$, $p = .32$; however, the more that participants imbued their confronter with growth mindset beliefs about prejudice the more positively they rated the confronter on the favourability subscale of the backlash measure, $r = .42$, $p < .001$.

Table 1

$N = 159$ who could recall being confronted					
	M	*SD*	1	2	3
1. Perceived Mindset of the Confronter	3.9	0.87	-	-	-
2. Backlash-Complainer	2.94	1.25	-0.24*	-	-
3. Backlash-Favourability	5.35	1.16	0.34*	-0.61*	-
$N = 74$, free response confirmed by coder					

	M	*SD*	1	2	3
1. Perceived Mindset of the Confronter	3.97	0.89	-	-	-
2. Backlash-Complainer	2.64	1.1	-0.119	-	-
3. Backlash-Favourability	5.71	0.95	0.42*	-0.51*	-

*Table 1. Means, standard deviations, and correlations between variables for Study 3, top panel for the full analysable sample (N = 159) and bottom panel for the sub-sample which coders could confirm described an incident involving bias (N = 74, based on the participants' free response text). *p<.01.*

Discussion

The results of Study 3 offer real-world evidence that the more people who are confronted imbue their confronters with a growth mindset, the less backlash they exhibit, both reducing their ratings of the confronter as a complainer and increasing their favourability. While the results were not as robust for the complainer component of backlash when only analysing data for participants who provided full free responses, the results from the smaller sample support a comparable pattern on the favourability subscale of the backlash measure.

Study 4

Study 3 offers suggestive correlational evidence that viewing a confronter as holding a growth mindset suppresses backlash, from real world expressions of bias. In Study 4, we wanted to investigate the relationship between confrontation, perceived growth mindset, and backlash with greater control and therefore we returned to a scenario methodology. Study 4 was a 2 condition (confrontation vs. no confrontation) between-subjects design. We predicted that the confrontation condition would yield more backlash than the no confrontation condition (replicating past work); however, we also expected the confrontation condition would yield more growth mindset perceptions than the no confrontation condition (replicating our earlier studies). Thus, we predicted a suppressor mediation effect

(MacKinnon et al., 2000),[1] such that although confrontation (X) increases backlash (Y), this backlash would be weakened to the extent that there was a countervailing perception that the confronter endorses more growth mindset beliefs (M). In this way, backlash would be suppressed by growth mindset perceptions. This study also explored a behavioural intention that follows from backlash—people's willingness to interact with the confronter in the future.

Method

Participants

We again wanted to achieve a minimum of $N = 75$ per cell, even with a priori exclusions, so we posted a study requesting 200 people on MTurk with a qualification questionnaire that asked for race and gender, with an exclusion only on gender that routed men into the study and women out. Participants were 199 men, $M_{Age} = 36.2$, $SD = 12.23$, who self-identified their race/ethnicity as: 27 African American, 125 European American/White, 17 Latinx American, 16 Asian American, 8 Native American, 3 African American-European American/White, and 1 self-described "human." Following the a priori exclusion criteria, 5 participants were excluded for incorrectly answering either or both of the attention and manipulation check questions (3 from the no confrontation condition, 2 from the confrontation condition). A sensitivity power analysis for a two-tailed independent samples t-test conducted post-hoc using G*power indicates that $N = 194$ would detect an effect size $d = .40$ with 80% power and alpha = .05.

Procedure

---

[1] Note that in traditional mediation (Hayes, 2018), one or more mediating variables (*M*) *strengthen* the relationship between a predictor variable (*X*) and an outcome variable (*Y*). When the indirect effect (*ab*) linking X to Y through M is statistically controlled, traditional mediation expects a reduction in the magnitude of the direct effect linking X and Y (path *c'*) relative to the total effect (path *c*). In contrast to a traditional mediator effect, a suppressor mediator *weakens* (rather than strengthens) the relationship between a predictor variable (*X*) and an outcome variable (*Y*). So, when the indirect effect (*ab*) is statistically controlled in the context of a suppressor mediator, we expect to see an *increase* in the magnitude of the direct effect (path *c'*) relative to the total effect (path *c*; see MacKinnon et al., 2000). Following the logic of suppression mediation, we predicted that to the extent that confrontation yielded growth mindset perceptions, it would elicit less backlash.

After providing informed consent, all participants read the scenario described in Study 1, with the gender biased comment about hiring for diversity purposes, but all participants imagined themselves in the role of perpetrator (i.e., as actors)—communicating the biased statement to Emily. After reading the scenario, participants were randomly assigned to either the confrontation condition ("After a moment of silence, Emily confronts you about your statement. Emily clearly expresses her disagreement with your comment on diversity hiring in a calm but firm voice") or no confrontation condition ("After a moment of silence, Emily does not respond to your statement. She gets her coffee and says she needs to head back to her desk"). Participants then completed the following measures.

Backlash. Participants completed the same, field-standard backlash measure (Kaiser & Miller, 2001) used in Study 3, rating Emily on the two subscales of backlash described above (complainer $\alpha = .95$, favourability $\alpha = .97$).

Perceived Mindset of the Confronter. Participants completed the measure of their perceptions of Emily's mindset beliefs about the malleability of prejudice described in Study 1 ($\alpha = .88$).

Openness to Future Interactions. Participants rated how much they would "look forward to interacting with Emily" and "avoid future interactions with Emily at all costs" (reverse-scored) on a 6-point scale from "*Strongly Disagree*" to "*Strongly Agree*" ($r = .72$, $p < .001$).

Attention and Manipulation Check. Participants completed one multiple choice item assessing their ability to recall who made a gender biased comment (themself or Emily) and one true/false item indicating whether they were in the confrontation or no confrontation condition.

<div align="center">**Results**</div>

To test the hypotheses, we conducted independent samples t-tests by condition and evaluated an indirect effects model using the SPSS v.27 PROCESS macro (Model 4).

Backlash. As expected, participants in the confrontation condition rated Emily as higher on the complainer subscale of the backlash measure ($M = 3.45$, $SD = 1.75$) than participants in the no confrontation condition ($M = 2.81$, $SD = 1.48$), $t(185.68) = -2.72$, $p = .007$, $d = 0.39$, and participants in the confrontation condition rated Emily lower on the favourability subscale of the backlash measure ($M = 4.26$, $SD = 1.13$) than participants in the no confrontation condition ($M = 4.57$, $SD = 1.04$), $t(189) = 1.97$, $p = .05$, $d = 0.28$.

Perceived Mindset of the Confronter. As predicted, participants in the confrontation condition rated Emily as endorsing more growth mindset beliefs ($M = 3.77$, $SD = 0.97$) than did participants in the no confrontation condition ($M = 3.40$, $SD = 0.77$), $t(189) = -2.93$, $p = .004$, $d = 0.42$. As hypothesized, perceived confronter mindset significantly correlated with backlash. The more participants rated Emily as having a growth mindset, the less they saw her as a complainer, $r = -.42$, $p < .001$, and the more they rated her favourably, $r = .33$, $p < .001$.

Process Model. We next examined whether growth mindset beliefs would operate as a suppressor mediator, weakening the impact of confrontation on backlash. We used Process Model 4 with 10,000 iterations to test whether an indirect effect of confrontation condition (X) through perceived mindset beliefs (M) would emerge on the complainer ratings (Y). This model was supported, *indirect effect* = -.016, *bootstrap SE* = 0.06, 95% *CI* (-0.28, -0.05). The direct effect of confrontation condition on complainer ratings remained significant as well, *B* = 0.48, *SE* = 0.11, $t(189) = 4.57$, $p < .001$, 95% *CI* (0.28, 0.69), and in line with suppression mediation this direct effect was stronger than the total effect, *B* = 0.32, *SE* = 0.12, $t(189) = 2.72$, $p = .007$, 95% *CI* (0.09, 0.55). A comparable pattern emerged on favourability, with a significant indirect effect, *indirect effect* = 0.09, *bootstrap SE* = 0.03, 95% *CI* (0.03, 0.16),

and a significant direct effect, B = -0.24, SE = 0.07, $t(189)$ = -3.22, p = .002, 95% CI (-0.39, -0.09) that was larger than the total effect, B = -0.15, SE = 0.08, $t(189)$ = -1.97, p = .05, 95% CI (-0.31, 0.0002). In sum, although confronting (vs. remaining silent) led to increased backlash toward Emily—in the form of higher complainer ratings and lower favourability ratings—these backlash effects were attenuated by perceiving her to endorse growth mindset beliefs.

The following analyses are exploratory.

Openness to Future Interactions. Participants in the confrontation (vs. no confrontation) condition reported less desire to interact with Emily in the future (M = 3.39, SD = 1.33 and M = 3.86, SD = 1.18, respectively), $t(189)$ = 2.54, p = .01, d = 0.37. Given this pattern, we again used Process Model 4 with 10,000 iterations to test whether an indirect effect of confrontation condition (X) through mindset perceptions (M) would emerge on the openness to future interactions (Y), which was supported, *indirect effect* = 0.10, *bootstrap SE* = 0.04, 95% CI (0.03, 0.18). Consistent with suppression mediation, the direct effect of confrontation condition on openness to future interactions ratings was significant, B = -0.33, SE = 0.09, $t(188)$ = -3.74, p < .001, 95% CI (-0.50, -0.15), and larger than the total effect, B = -0.23, SE = 0.09, $t(188)$ = -2.54, p = .01, 95% CI (-0.41, -0.05). So, although people were less open to interacting with Emily after she confronted bias, perceiving Emily to endorse growth mindset beliefs attenuated this negative outcome.

Given this, we further explored whether the mindset perceptions and complainer and favourability ratings might serially mediate this effect on willingness to interact. The complainer subscale of the backlash measure correlated significantly with openness to future interactions, r = -.59, p < .001, as did the favourability subscale, r = .64, p < .001. We used Process Model 6 with 10,000 iterations to test whether an indirect effect of confrontation condition (X) through perceived mindset beliefs (M_1) and complainer ratings (M_2) would

emerge on the openness to future interactions. This serially mediated pathway was supported, *indirect effect* = 0.07, *bootstrap SE* = 0.03, 95% *CI* (0.02, 0.12). When then used Process Model 6 with 10,000 iterations to test whether an indirect effect of confrontation condition (X) through perceived mindset beliefs (M_1) and favourability ratings (M_2) would emerge on the openness to future interactions and this mediational pathway was also supported, *indirect effect* = 0.06, *bootstrap SE* = 0.02, 95% *CI* (0.02, 0.11). Taken together, although people were less interested in interacting with Emily when she confronted bias (vs. remained silent), to the extent Emily was afforded growth mindset beliefs, she was viewed to be less of a complainer and more favourable, which prompted greater openness to interacting with Emily in the future.

Discussion

In this study, we manipulated whether participants imagined being confronted (or not) after expressing gender bias. Then we measured their backlash against the confronter, their perceptions of the confronter's mindset beliefs about the malleability of prejudice, and their openness to interacting with the confronter in the future. The results overall support the hypotheses. We replicated the classic backlash effect, showing simultaneously both higher ratings of the confronter as a complainer and lower ratings of her favourability. A robust and hypothesized suppression effect also emerged: participants in the confrontation condition imbued the confronter with more growth mindset beliefs about prejudice and, to the extent that growth mindset attributions were made, participants exhibited less backlash toward the confronter. As we report in detail in the SOM, analyses controlling for offensiveness and likelihood of making the statement offer consistent support for the Process models although the condition main effects are weakened. We can only speculate as to why controlling for offensiveness and likelihood of making the statement have these uneven effects in this study—which themselves are not consistent across studies—but it may suggest that future

investigations ought to systematically vary the believability and level of offensiveness of a biased comment to fully explore the role of these perceptions.

<div align="center">**Study 5**</div>

We next sought to extend the previous two studies by offering further triangulated evidence for causality. Here, we manipulated the theorized mediator (growth mindset perceptions) and measured backlash as the outcome (Spencer et al., 2005). We conducted four 2 (Confrontation Condition: confrontation vs. no confrontation) × 2 (Mindset Condition: growth vs. fixed) between-subjects experiments to further explore the dynamics of backlash suppression. Because the methods, manipulation, and measures were so similar across Studies 5a, 5b, 5c, and 5d, and because larger samples offer more reliability and robustness of interpretation, we present the results as a single large-sample meta-analysis (see SOM for study-by-study reporting; Goh et al., 2016).

In the initial studies (5a, 5b), we predicted an interaction would emerge such that the growth (vs. fixed) mindset message would fully mitigate backlash against the confronter (vs. no confrontation). This hypothesis was not supported in Study 5a or 5b, so for Study 5c and Study 5d we no longer predicted that a growth mindset would fully suppress backlash. Instead, and for the meta-analysis, we hypothesized the emergence of main effects of both confrontation condition and mindset condition, but no interaction. That is, we expected participants would show more backlash against Emily—perceiving her simultaneously as more of a complainer and less favourably when she confronted prejudice (vs. when she did not). At the same time, however, when Emily was described as endorsing more growth (vs. fixed) mindset beliefs, we expected that participants would attenuate their backlash against Emily—perceiving her to be relatively less of a complainer and perceiving her relatively more favourably. This possibility would be consistent with our emerging theorizing that to the extent that a person is viewed as having a growth mindset, it can mitigate—albeit not

fully eliminate—backlash. In this study, we invoked mindsets about the person as a whole to see whether the effect of confrontation would be extended to personality mindsets which have been more closely tied to prejudice confrontation in past research (Rattan & Dweck, 2010, 2018).

Method

Participants

The target sample sizes, all determined a priori, for Study 5a was 400 men and women on MTurk, 5b was 400 men on Prolific Academic, 5c was 800 men on MTurk (though data collection was closed at 650 due to a lack of further sign ups), and 5d was 800 men and women on MTurk. Because the studies included disproportionately more men than women, for the meta-analysis we retained men only (for completeness, women's analyses are reported in the SOM). We included only those who passed the attention and manipulation checks. Participants were 1,622 men, $M_{Age} = 33.71$, $SD = 11.08$, who self-identified their race/ethnicity as: 103 African American, 1129 European American/White, 94 Latinx American, 187 Asian American, 23 Native American, 2 Middle Eastern American, 69 participants who indicated biracial or multiracial identities, and 14 who identified as other. A sensitivity power analysis for an ANOVA (fixed effects, special, main effects and interactions) conducted post-hoc using G*power indicates that $N = 1622$ would detect an effect size $f = .07$ with 80% power and alpha = .05. Exclusions ($N = 1060$, who did not self-identify as men, who answered the manipulation or reading check questions incorrectly) were equal across conditions, χ^2 (2, 1058) = 2.03, $p = .154$.

Procedure

After providing informed consent, participants in all studies read about Emily, who was again described as working in a large professional services organization, loving her job, and being a high performer. In the course of this description, participants learned about a

recent conversation Emily had, in which a friend asked her how much she thought people could change. Participants were randomly assigned to the mindset condition, either reading that, "Emily replied, 'Someone's personality is a part of them that they can't change very much'" (fixed mindset condition) or that, "Emily replied, 'No matter who somebody is and how they act, they can always change their ways'" (growth mindset condition). All participants then read a scenario of workplace gender bias used in the earlier studies and were randomly assigned to read either the confrontation or no confrontation condition, as described in the earlier studies. All versions of the study put participants in the observer perspective, such that the comment occurred between Emily and John. Studies 5a, 5b, and 5c included the emotional women managers comment while Study 5d included the comment about hiring for diversity reasons. After the scenario, participants completed the backlash measure described in Study 3 (complainer α = .92; favourability α = .97). They also completed attention and manipulation check questions, specifically a true/false item testing their knowledge of Emily's mindset and three multiple choice items assessing their knowledge of the facts of the scenario and the confrontation condition they had been assigned to. Participants then completed additional measures assessing how offensive they thought the comment was and their agreement with it (see SOM for details) and a standard demographics measure.

Results

Consistent with hypotheses, participants in the confrontation condition (M = 2.24, SD = 1.23) rated Emily as more of a complainer than did participants in the no confrontation condition (M = 1.91, SD = 1.03), $F(1, 1621)$ = 35.53, $p < .001$, n_p^2 = .02. Further, participants in the growth mindset condition (M = 1.98, SD = 1.11) rated Emily lower on the complainer subscale of the backlash measure than participants in the fixed mindset condition (M = 2.16, SD = 1.18), $F(1, 1621)$ = 11.57, p = .001, n_p^2 = .007. The interaction was not significant, $F(1, 1621)$ = 0.19, p = .66, $n_p^2 < .001$, meaning an explicit endorsement of a growth mindset led to

lower complainer perceptions in both conditions including, most critically, the confrontation condition (see Figure 2a). Thus, although confronting led to greater perceptions that a person was a complainer, these complainer perceptions were weakened (though not removed) when it was clear that the confronter endorsed more growth mindset beliefs.

For the favourability subscale of backlash, participants in the growth mindset condition ($M = 5.50$, $SD = 1.00$) rated Emily more favourably than participants in the fixed mindset condition, ($M = 5.22$, $SD = 1.00$), $F(1, 1621) = 30.93$, $p < .001$, $n_p^2 = .019$. Confrontation condition had no effect on favourability ratings, $F(1, 1621) = 0.71$, $p = .40$, $n_p^2 < .001$, and there was no interaction, $F(1, 1621) = 0.36$, $p = .55$, $n_p^2 < .001$ (see Figure 2b). In other words, Emily's growth (compared to fixed) mindset led participants to perceive her more favourably, and this effect held regardless of whether she confronted John. Together, the results for the complainer and favourability subscales of the backlash measure consistently indicate that the growth mindset manipulation led to more favourable impressions and indeed reduced (but did not remove) backlash by attenuating the extent to which confronters were viewed as complainers.

Conducting the analyses above controlling for the type of comment used (emotional women managers or diversity hiring) and the recruitment platform (MTurk or Prolific Academic) yielded similar results and leave the conclusions of this study unchanged.

Figure 2

(a) Perceptions that confronter   (b) Perceptions that confronter
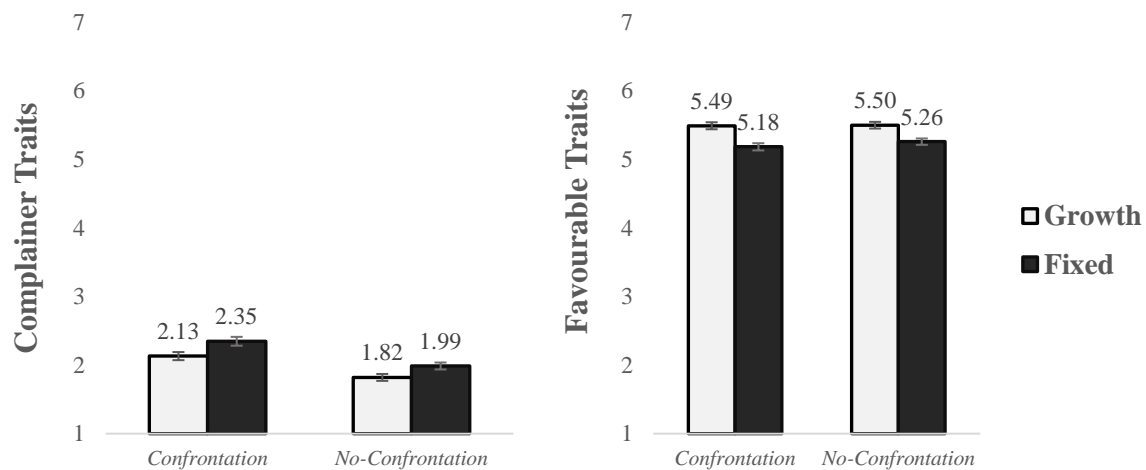   is a complainer   is favourable

Figure 2. Summary of means (and standard errors) for perceptions of confronters collapsed across Studies 5a-5d: complainer traits (panel a), favourable traits (panel b).

Discussion

Study 5 offers a large-sample investigation of whether a brief manipulation of a confronter's mindset can affect backlash against the confronter. While backlash was still observed, when the confronter communicated growth (vs. fixed) mindset beliefs, it significantly attenuated people's perceptions of her as a complainer and increased their favourability. Because the sample size of this study was so large, we are more confident in the estimate of the true effect size of perceived growth mindset beliefs on backlash. We see that this light touch manipulation had a small, but statistically significant effect on backlash. As we report in detail in the SOM study-by-study analyses, the condition manipulations did not consistently yield differences in either participants' perceptions of the offensiveness of the statement or their agreement with it.

Study 6

Next, we wanted to assess whether a stronger manipulation would yield a more substantial effect. We also sought to move away from scenario methods which lack real-world validity. Thus, in the final study, we engaged participants in a lab task in which they exhibited bias by expressing racial stereotypes. A confederate confronted the participant,

mentioning their own mindset beliefs, in this context about the malleability of prejudice, in the course of confrontation. By embedding the mindset in the confrontation, we hoped to increase the strength of the mindset manipulation and its potential influence on backlash toward a confronter. This study also shifts the burden of confrontation from targets of prejudice to allies, given that the confederate was a White person confronting anti-Black bias.

We employed a 3 condition (growth mindset beliefs; fixed mindset beliefs; vs. no mindset beliefs, confrontation-only control) between-subjects design.

Method

Participants

We initially aimed to recruit a minimum of $N = 240$ (80 per cell) student participants for this in-person laboratory experiment; however, we were only able to recruit 203 participants and because we were unable to restart the study we did not achieve our recruitment goal: 132 women, 58 men, 13 who chose not to disclose their gender identity (growth mindset condition = 65; fixed mindset condition = 71, confrontation-only control condition = 67). Since participants in this study were confronted for expressing anti-Black racial bias, we only retained participants who identified (at least partially) as European American/White ($N = 189$; growth mindset condition = 65; fixed mindset condition = 61, confrontation-only control condition = 63). Even amongst this group, there was some racial diversity. Students self-identified as: 178 European American/White, 4 European American/White- Latinx American, 3 European American/White-Native American, 2 European American/White-Asian American, 1 European American/White-African American-Latinx American, 1 European American/White-Other. As preregistered, we excluded people if more than 75% of their data was missing ($N = 8$), resulting in a final sample of 181 participants (growth mindset condition = 62; fixed mindset condition = 59, confrontation-only control condition = 60). A sensitivity power analysis for an ANOVA (fixed effects,

special, main effects and interactions) conducted post-hoc using G*power indicates that $N = 181$ would detect an effect size $f = .23$ with 80% power and alpha = .05.

Procedure

Upon arrival to the lab, participants were greeted by an experimenter and seated in a testing room. A few minutes later, a trained confederate (posing as a participant) arrived and was also seated in the same testing room. The study ostensibly examined differences between 'individual work' and 'group work.' Participants were told by the experimenter that they would complete some tasks individually and others as a pair. The experimenter then obtained informed consent.

First, the participant and confederate worked together on the Photograph-Sentence Pairs task established in past prejudice confrontation research to elicit stereotyping (e.g., Burns & Monteith, 2018; Czopp et al., 2006). The experimenter casually handed the confederate (but not the participant) a paper packet and a pen. This was done so that during this first task the participant would familiarize themselves with the confederate's handwriting. This would bolster the credibility of the handwritten confrontation message the confederate would later provide. In the Photograph-Sentence Pairs task, participants were shown photographs of individuals from a diverse set of racial and gender groups. Underneath each photograph was a short sentence stem that the participant and confederate would work together to complete. For example, in one trial participants saw a picture of a middle-aged White man paired with the sentence stem: "This person works with paint—this person is likely a _____." The participant and confederate would then discuss and find a mutually agreed upon way to complete the sentence stem (e.g., "…this person is likely an *artist*" or "…this person is likely a *house painter*"). Of the 18 total trials for this task, 3 were critical trials, where participants saw a photograph of a Black person with a sentence stem designed to invoke a racial stereotype. For example, one of the critical trials paired a

photograph of a Black man with the sentence stem: "This person can be found behind bars.

This person is likely a _____." This sentence stem is meant to invoke Black

criminality stereotypes, so stereotype-consistent answers would be "prisoner" or "criminal";

but there are alternative, stereotype-inconsistent answers that also logically complete the

sentence stem, like "bartender" or "coffee barista." For these critical trials, confederates were

instructed to let the participant offer an answer to the sentence stem first and to go along with

whatever the participant said. Throughout this task, the confederate was blind to condition.

As expected, most participants (84.7%) submitted at least one anti-Black stereotype when

completing this task. Anecdotally according to the confederates, even the small percent who

did not submit an anti-Black stereotype initially stated one aloud before changing to a less

stereotypical answer.

After the Photograph-Sentence Pairs task, the participant and confederate were

ushered into separate testing rooms to complete several surveys. In their first survey set,

participants were asked to write a short note to their partner (i.e., *What feedback, if any,*

would you like to give your partner about his/her work during the Photograph-Sentence

Pairs task?)—and, to the participant's knowledge, their partner was in another testing room

doing the same thing for them. Once finished, the experimenter took the participant's written

note indicating that they would be right back with the note from their partner.

When the experimenter returned, they gave the participant a handwritten note from

their partner (the confederate). Participants were randomly assigned to receive one of three

confrontation messages. In the growth mindset condition, the note read:

> "I know this task was a little weird, but I thought some of your answers on the photos of black
>
> people were a little prejudiced, and that bothered me. Like that one where you said the

[guy/girl] was a [stereotype used by participant, e.g., "criminal"].[2] I wanted to point it out

because I think people can work on these things and change how biased they are."

It explicitly indicated that they believed racial biases are malleable. By contrast, in the fixed

mindset condition, the note read:

"I know this task was a little weird, but I thought some of your answers on the photos of black

people were a little prejudiced, and that bothered me. Like that one where you said the

[guy/girl] was a [stereotype used by participant]. I wanted to point it out, but I don't think

people can do much to change how biased they are."

Here, the note explicitly indicated that their partner believed racial biases are unchangeable.

Finally, in the confrontation-only control condition, the partner note simply read:

"I know this task was a little weird, but I thought some of your answers on the photos of black

people were a little prejudiced, and that bothered me. Like that one where you said the

[guy/girl] was a [stereotype used by participant]."

It contained no specific information detailing the confronter's mindset beliefs about the

malleability of prejudice.

After reading the confrontation message from their partner, the participant provided

their immediate emotional reaction to the feedback and then completed the Photograph-

Sentence Pairs Task again, with a new set of 18 trials. Again, the task included 3 critical trials

where participants saw a photograph of a Black person and a stereotype eliciting sentence

stem. By completing the Photograph-Sentence Pairs Task twice (once before and once after

confrontation), we can compare whether participants behaviourally responded to the

confrontation by reducing their level of stereotyping or not, and whether any change varied

by condition.

Next, participants completed the following measures.

---

[2] For the minority of participants who did not provide any anti-Black stereotypes on the three critical trials of
the Photograph-Sentence Pairs task, the confederate omitted this line from their confrontation message.

Perceived Mindset of the Confronter. As a manipulation check, we assessed whether participants perceived their partner to endorse more growth mindset beliefs about prejudice. Participants indicated their agreement with two items: "My partner sees prejudice as a malleable trait that can be changed" and "My partner sees prejudice as a fixed trait that cannot be changed" (reverse-scored) on a 6-point scale from "*strongly disagree*" to "*strongly agree*," with higher numbers indicating that they perceived their partner to endorse more growth mindset beliefs (α = .919).

Backlash. We did not think the previously-used backlash measure would be appropriate for this study, as the complainer measure might cue to participants the true purpose of the study given how extreme some of the items are. Therefore, this study assessed backlash against the confronter using a measure of perceived warmth. To assess perceptions of the confronter's warmth, participants rated how much they perceived their partner to be "friendly," "likeable," "warm," and "approachable," and the extent to which they "would like to be friends with" their partner (α = .921, 5-items, 1 "*strongly disagree*" to 6 "*strongly agree*"). Less warmth indicated greater backlash as operationalized with this measure.

Openness to Future interactions. Next, to measure participants' willingness to interact with their partner again in the future, participants answered a single item about their willingness to complete another lab experiment with their partner (i.e., "If you revisited the lab, to what extent would you want to continue working with the partner you worked with today?", 1 "*not at all interested*" to 5 "*extremely interested*"). Additionally, participants responded to five items, including "If I had the choice, I wouldn't work with my partner again" and "I wouldn't want to hang out with my partner in the future", which were reverse-scored (α = .862, 5-items, 1 "*strongly disagree*" to 6 "*strongly agree*"). Our pre-registered plan was to analyse the single-item measure separately from the composite of the 5-item measure.

Attention Check. Participants answered 1 multiple choice attention check question, which asked them to recall an accurate statement from the handwritten note written by their partner (out of 5 possible statements). Unexpectedly, attention check failure was quite high across all three conditions (growth mindset condition = 17 failures; fixed mindset condition = 13 failures, confrontation-only control condition = 25 failures). We suspect that these high failure rates were due to the wording of the attention check question. We asked students to report *exactly* what their partner wrote to them, but the response options provided were only *close* to what their partner actually wrote. For this reason, we chose to retain participants, regardless of attention check failure. This decision deviated from our preregistration plan.

Finally, participants completed a standard demographics form, were debriefed, and compensated. See SOM for additional measures .

Results

Perceived Mindset of the Confronter. As predicted, condition significantly impacted participants' perceptions of the confronter's mindset beliefs, $F(2, 177) = 61.22$, $p < .001$, $\eta_p^2 = .409$. Participants in the growth mindset condition perceived the confronter as endorsing more growth mindset beliefs about prejudice ($M = 5.55$, $SD = 1.44$) compared to participants in the fixed mindset condition ($M = 2.57$, $SD = 1.71$, $p < .001$, $d = 1.91$). Also, replicating the earlier studies, participants in the control condition—who were simply confronted by their partner, but not provided with explicit mindset belief information—also expected the confronter to endorse more growth mindset beliefs ($M = 4.88$, $SD = 1.46$) compared to participants the fixed mindset condition ($p < .001$, $d = 1.47$), suggesting again that confrontation itself is a growth mindset cue. Still, these data suggest that confronters who were explicit about their growth mindset beliefs were rated as marginally more growth-mindset ($p = .054$, $d = 0.46$) compared to the confronters in the confrontation-only control

condition. These findings indicate that our manipulation of confronter mindset beliefs was successful.

Backlash. Next, we tested whether mindset condition influenced participants' perceptions of the confronter's warmth. Directionally consistent with our predictions, we found a marginal but nonsignificant main effect of condition, $F(2, 177) = 2.47$, $p = .088$, $\eta_p^2 = .027$. Supporting our pre-registered directional hypotheses between conditions, participants in the growth mindset condition perceived the confronter who explicitly espoused growth mindset beliefs to be significantly warmer ($M = 4.40$, $SD = 0.91$) than participants in the fixed mindset condition ($M = 4.03$, $SD = 1.17$, $p = .05$, $d = 0.36$) and warmer than participants in the confrontation-only control condition ($M = 4.04$, $SD = 1.02$, $p = .062$, $d = 0.37$), though this was marginal but not significant, suggesting that the explicit growth mindset message shaped warmth perceptions. Perceptions of the confronter's warmth did not differ between the fixed mindset and confrontation-only control conditions ($p = .930$, $d = 0.02$).

Openness to Future interactions. There were no condition differences on participants' willingness to interact with the confronter in the future (for neither the single item measure, $F(2, 178) = 0.10$, $p = .902$, $n_p^2 = .001$, nor the composite measure, $F(2, 178) = 0.20$, $p = .823$, $n_p^2 = .002$). However, to test our full, pre-registered, hypothesized model we used Process Model 4 with 10,000 iterations to explore whether an indirect effect of mindset condition (X) through backlash (M) would emerge on openness to future interactions (Y).[3] This model was supported. First, we examined the single-item willingness to interact measure as the outcome. Significant indirect effects emerged when comparing the fixed mindset condition to the growth mindset condition, *indirect effect = 0.26, bootstrap SE = 0.13, 95%*

---

[3] Because mindset condition has three levels, we used the multicategorical indicator coding system, built into Process (Hayes & Montoya, 2017).

CI (0.00, 0.53) and when comparing the confrontation-only control condition to the growth mindset condition, *indirect effect* = 0.25, *bootstrap SE* = .13, 95% *CI* (0.004, .50). The indirect effect was not significant when comparing the fixed and confrontation-only control conditions, *indirect effect* = 0.01, *bootstrap SE* = 0.14, 95% *CI* (-0.27, 0.29).

Second, we examined the 5-item openness to interaction composite measure as the outcome. Again, significant indirect effects emerged when comparing the fixed to the growth mindset condition, *indirect effect* = 0.32, *bootstrap SE* = 0.16, 95% *CI* (0.00, 0.65) and when comparing the confrontation-only control to the growth mindset condition, *indirect effect* = 0.30, *bootstrap SE* = 0.15, 95% *CI* (0.01, 0.60). As before, the indirect effect was not significant when comparing the fixed and confrontation-only control conditions, *indirect effect* = 0.01, *bootstrap SE* = 0.17, 95% *CI* (-0.32, 0.36). Together, these findings demonstrate that growth (vs. fixed and vs. control) mindset confrontations attenuate backlash which in turn is associated with greater openness to interact with the confronter in the future.

Stereotyping. There were no differences in participants' racial stereotyping behaviours by mindset condition, $F(1, 178) = 0.22$, $p = .803$, $n_p^2 = .002$. Across all three mindset conditions, and consistent with past confrontation research (e.g., Czopp et al., 2006), participants tended to stereotype less after being confronted ($M = 0.69$, $SD = 0.88$), compared to before being confronted ($M = 1.54$, $SD = 0.95$), $F(1, 178) = 125.01$, $p < .001$, $n_p^2 = .413$. Whereas 62.98% of participants used fewer anti-Black racial stereotypes after being confronted (compared to before), a much smaller proportion of participants stereotyped to the same degree (16.58%) or stereotyped more (8.29%) after confrontation. The remaining 12.15% did not use anti-Black racial stereotypes at all (neither before nor after confrontation).

Discussion

In Study 6, we moved away from the scenario methods employed in earlier studies to a method with more real-world validity. Participants were confronted for using anti-Black racial stereotypes, and the ally confronter communicated their mindset beliefs within the confrontation itself. Participants who learned their confronter endorsed more growth (vs. fixed) mindset beliefs about prejudice exhibited less backlash by perceiving the confronter more positively (more warmly). Further, through higher perceptions of confronter's warmth, the growth mindset (vs. fixed and vs. control) confrontation predicted increased openness to interacting with the confronter again in the future. Lastly, we tested the possibility that a growth mindset (vs. fixed and vs. control) confrontation, by virtue of communicating greater warmth, might undermine the effectiveness of the confrontation, in terms of the degree of reduced stereotyping that followed. We found no such evidence. Instead, Study 6 showed an equal (and significant) degree of reduced stereotyping in the 2nd round of the task, confirming that a growth (vs. fixed) mindset confrontation can both be effective (i.e., reduce stereotyping) and suppress interpersonal backlash (as captured by warmth and future interaction intentions). For a non-pre-registered, exploratory analysis investigating our specific theoretical mechanism on an item that incidentally captures this construct (i.e., the idea that the growth mindset (vs. fixed, vs. control) confrontation communicates more that the confronter believes the confronted party will change), please see SOM.

General Discussion

Across 10 studies, we investigated two core questions. First, would people perceive someone who confronts (vs. remains silent to) a biased statement as holding a more growth mindset about prejudice? Second, would these growth mindset perceptions suppress the degree to which people exhibit backlash against confronters? We investigated these questions using scenarios with two different gender-biased comments (Studies 1-2b, 4-5), naturalistic confrontations of national, race, and gender stereotypes reported retrospectively (Study 3),

and an in-person laboratory experiment of actual confrontations of racial bias (Study 6). Both correlational and experimental methods yielded support for our core hypotheses: People spontaneously imbued someone who confronts a biased comment with more growth mindset beliefs about prejudice and bias (Studies 1-2b, 4, 6), and these growth mindset perceptions suppress backlash, as assessed by classic measures of backlash in interpersonal perceptions (Studies 4-5) and judgments of interpersonal warmth and willingness to interact again in the future (Study 6). Across studies, the patterns were consistent both when the confronter was a target of the biased behaviour (Studies 1-5), and when they were an ally who did not belong to the targeted group (Study 6). Moreover, this pattern emerged both when the bias was expressed by someone else (Studies 1-2b, 5), or when the participant made the biased comment (Studies 2a-4, 6) and in both correlational studies (Study 3-4) and when a growth mindset (about personality, Study 5, about prejudice, Study 6) was manipulated, confirming causality. Moreover, confrontations with embedded growth mindset messages were equally effective in reducing people's stereotyping behaviour (Study 6). In sum, the current research reveals that the act of confronting biased behaviour signals a greater growth mindset about prejudice to perceivers and, in so doing, may serve to attenuate the level of backlash people exhibit against confronters.

Theoretical Implications

The current research advances our theoretical understanding of mindsets and the nascent study of people's perceptions of others' mindset beliefs. Early investigations of people's mindset perceptions established that people are motivated to perceive others' mindsets, and demonstrate that their mindset perceptions meaningfully shape their own motivation, experiences, and outcomes (Fuesting et al., 2019; Good et al., 2012; LaCossse et al., 2021; Muenks et al., 2020; Rattan et al., 2018). Our work is some of the first to open the black box examining how people come to form their perceptions of others' mindset beliefs,

particularly within the domain of prejudice (vs. ability/intelligence). We offer evidence that specific behaviours or actions—in this case, confronting biased behaviour—can cause people to view the confronter as endorsing more growth, compared to fixed, mindset beliefs.

This novel link between bias confrontation and perceived mindset beliefs opens up new avenues for scholarship investigating other types of behaviours that may signal people's mindset beliefs. In addition, this research advances literature on the consequences of perceived mindset beyond the self-relevant outcomes explored in educational contexts (e.g., student's own feelings, motivation, and performance, Fuesting et al., 2019; Good et al., 2012; Muenks et al., 2020; Rattan et al., 2018), toward understanding how perceived mindset shapes people's treatment of others—in this case the degree of backlash they exhibit toward targets and allies who confront. This work raises the possibility that we are only starting to fully understand the array of outcomes shaped by mindset perceptions (Rattan & Ozgumus, 2019).

Our research also contributes to the study of bias confrontation and intergroup relations. While decades of research emphasize the importance of intergroup contact and its profound potential to improve intergroup relations (Allport, 1954; Pettigrew & Tropp, 2000), we know much less about how intergroup contact can recover when acute instances of bias arise in the course of these everyday interactions. To date, the act of bias confrontation has been considered a moment of intergroup conflict. Our findings add to past work (Rattan & Dweck, 2010; 2018; Rattan, 2019) by suggesting that bias confrontation can instead be construed as a moment of intergroup learning, insofar as it engages a growth mindset. Of course, this may not apply to all confrontations depending upon their content and tone, which we return to below and which future research should explore. Secondly, aside from notable research which shows that questions which affirm someone (e.g., asking about times when the person has been fair, or creative) can reduce the backlash they exhibit to someone who

confronts bias (Stone et al., 2011), there are few evidence-based recommendations to inform scholars and practitioners about how to reduce the backlash that confronters typically face. Our research contributes to the study of confrontation backlash by highlighting that communicating one's growth mindset beliefs in advance of confronting, or embedding one's growth mindset beliefs into the confrontation content itself, may be an effective means to suppress the backlash confronters often receive, offering evidence from different methods (e.g., scenarios, retrospective naturalistic confrontations, in-person confrontation) and from multiple perspectives (e.g., examining those confronted about their own bias and observers of someone else being confronted about a biased comment). Notably, our studies simultaneously show how pernicious and persistent backlash is, indicating that future research might investigate what else can be added alongside growth mindset beliefs to more substantially reduce backlash.

Limitations and Future Directions

While we extensively examined our core research questions across 10 studies, we of course acknowledge that our investigation was not comprehensive in assessing mindset perceptions and backlash against all the social groups that experience bias in society today. Our scenario studies examined the perceived mindset beliefs of a woman, whose name (Emily) likely evoked a White woman prototype. Our goal for future work would be to investigate a more diverse array of identities on the confronter side, to test whether confronters (e.g., Black women, LGBTQ+ individuals, men allies, disabled individuals) may be differentially afforded growth mindset perceptions. While our studies of actual interpersonal interactions included high levels of national diversity (Study 3) and ally confrontation (Study 6), these studies still cannot speak to this important question because confronters' identities were not systematically varied. It would also be fascinating to investigate whether the degree of growth mindset perceptions afforded to a confronter varies

based on the type of bias they are speaking out against. For example, would people spontaneously afford equal growth mindset perceptions to someone who confronts ageism, given past work showing that people hold greater fixed mindset beliefs about the potential of older adults to change (Neel & Lassetter, 2015)? It is possible that not all confrontation types (across confrontation content, or even tone as we discuss below) would spontaneously evoke a growth mindset. We look forward to future research investigating these questions.

We must also note that this research was largely conducted in the United States, across a set of years in which the #MeToo and #BlackLivesMatter movements emerged (2017-2020). Speaking out to confront a biased comment may be a behaviour that is particularly valued in North American and some Western European, individualist contexts which place less emphasis on concerns about saving face (Lee et al., 2012), suggesting that future research should test whether our findings are limited to cultural contexts that share these values (although, our Study 3 suggests some international generalizability for the patterns observed). Additionally, these social movements could have created a wider discourse that demands change through speaking out about bias and harassment and self-education, which could be how confrontation came to be associated with growth mindset perceptions. We consider this a fascinating question and comparative research that examines contexts in which these movements are more nascent could offer insight into this question, and thus illuminate some broader cultural boundary conditions that may exist on the phenomenon we have documented. In this context, it may be the case that a confrontation which includes a concrete growth mindset message (e.g., perhaps expanded text based on our Study 6), might foster a growth mindset in the person being confronted, encouraging them to see their own prejudices and biases as more malleable. If this were to be the case, growth mindset confrontations might also represent a new avenue for mindset interventions, in the context of bias and diversity. We look forward to future research that can test these

possibilities in the context of real-life interactions, where the impact of a growth mindset message during confrontation may be most meaningful. Furthering these ideas, research might test whether the construal level (Trope & Lieberman, 2010) of the confrontation (i.e., whether the confrontation is described generally or in specific quotations) might shape the effectiveness of a growth mindset message in suppressing backlash – it may be that when third parties hear specific quotations (as we used in Study 6) they afford stronger or weaker growth mindset perceptions than when, for example through an office gossip chain, a story is relayed in general (as in our earlier studies).

Notably, across our studies the confrontations being observed were characterized by a calm tone and, in some cases, may have minimized the perpetrator's transgression (e.g., "your answers… were *a little* prejudiced" in Study 6). Yet, past research has shown that people exhibit meaningfully different responses to confrontations depending upon whether they are communicated in firm, hostile, and angry tones, versus yielding, calm, and neutral tones (Czopp et al., 2006). Would confrontations across different affective tones evoke similar levels of growth mindset perceptions, and would any level of growth mindset perceptions evoked within these confrontations similarly suppress backlash? To date, no research to our knowledge has studied how emotional signals intersect with growth versus fixed mindset perceptions, meaning that this question provides a rich context for further theoretical development in the study of mindsets. We look forward to future research which will evaluate whether high arousal negative emotions (e.g., anger, hostility) or low arousal negative emotions (e.g., despair, shame) expressed in a confrontation differentially shape mindset perceptions and subsequently the suppression of backlash. It may seem natural to think that infusing a confrontation with anger, for example, might reduce growth mindset perceptions. However, anger is a high arousal positive emotion associated with agency, and in the context of prejudice confrontations the growth mindset is also associated with greater

motivation and engagement (Rattan & Dweck, 2010). This overlap in agency raises the possibility, perhaps remote, that anger could possibly *add* to the perception of growth mindsets following a confrontation. We see the study of low arousal negative emotions as equally promising. For example, perhaps a confrontation infused with despair (e.g., at the state of bias in the world) would evoke more sympathy and thus less backlash. If this were to be the case, could the combination of despair and a fixed mindset statement within the confrontation protect confronters from backlash (though through a different psychological process than studied in the current work)? We eagerly look forward to future research that further develops a theory of mindsets and emotion displays.

Conclusion

People who confront the biased behaviour of others should not be denigrated for their engagement in anti-biased action, but too often they are (e.g., Czopp & Monteith, 2003; Rasinski & Czopp, 2010). Indeed, people are aware of this risk and this knowledge can discourage people from speaking out (Good et al., 2012). Our work takes a step toward identifying a psychological process that interrupts this confrontation-related backlash by illuminating a novel link between confrontation and perceived growth mindset beliefs on the part of the confronter. Moreover, this work offers practical suggestions to people who wish to speak out, despite the social costs they may face. We show that when confronters communicate their growth mindset beliefs explicitly and/or when they embed their beliefs about people's ability to change into the content of their confrontation, it attenuates backlash and thus promotes future contact. Indeed, taken together, we find that attributing growth mindset beliefs to confronters may be one way to effectively undercut backlash and promote openness to future interaction with those who challenge bias when they see it.

References

Adams, G. S. (2016). Asymmetries between victims' and transgressors' perspectives following interpersonal transgressions. *Social and Personality Psychology Compass, 10*(12), 722-735. https://doi.org/10.1111/spc3.12291.

Allport, G. W. (1954). *The nature of prejudice.* Addison-Wesley.

Ashburn-Nardo, L., Blanchar, J. C., Petersson, J., Morris, K. A., & Goodwin, S. A. (2014). Do You Say Something When It's Your Boss? The Role of Perpetrator Power in Prejudice Confrontation: Perpetrator Power and Prejudice Confrontation. *Journal of Social Issues*, *70*(4), 615–636. https://doi.org/10.1111/josi.12082

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.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Frontiers in human neuroscience, 6*, 117. https://doi.org/10.3389/fnhum.2012.00117.

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. https://psycnet.apa.org/doi/10.1037/0021-9010.79.6.993.

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%2F1368430218798041.

Cadieux, J., & Chasteen, A. L. (2015). You gay, bro? Social costs faced by male confronters of antigay prejudice. *Psychology of Sexual Orientation and Gender Diversity, 2*(4), 436. https://psycnet.apa.org/doi/10.1037/sgd0000134.

Canning, E. A., Murphy, M. C., Emerson, K. T., Chatman, J. A., Dweck, C. S., & Kray, L. J. (2020). Cultures of genius at work: Organizational mindsets predict cultural norms, trust, and commitment. *Personality and Social Psychology Bulletin, 46*(4), 626-642. https://doi.org/10.1177%2F0146167219872473.

Canning, E. A., Ozier, E., Williams, H. E., AlRasheed, R., & Murphy, M. C. (2021). Professors Who Signal a Fixed Mindset About Ability Undermine Women's Performance in STEM. *Social Psychological and Personality Science*, 19485506211030400. https://doi.org/10.1177/19485506211030398

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856. https://doi.org/10.1037/0022-3514.66.5.840

Carlston, D. E., Skowronski, J. J., & Sparks, C. (19960101). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420. https://doi.org/10.1037/0022-3514.69.3.429

Carr, P. B., Dweck, C. S., & Pauker, K. (2012). "Prejudiced" behavior without prejudice? Beliefs about the malleability of prejudice affect interracial interactions. *Journal of Personality and Social Psychology, 103*(3), 452. https://psycnet.apa.org/doi/10.1037/a0028849

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%2F0146167202250923.

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. https://psycnet.apa.org/doi/10.1037/0022-3514.90.5.784.

Drury, B. J., & Kaiser, C. R. (2014). Allies against sexism: The role of men in confronting sexism. *Journal of Social Issues, 70*(4), 637-652. https://doi.org/10.1111/josi.12083.

Dweck, C. S., & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science, 14*(3), 481-496. https://doi.org/10.1177%2F1745691618804166

Dweck, C. S. (2013). *Self-theories: Their role in motivation, personality, and development.* Psychology press. https://doi.org/10.4324/9781315783048.

Focella, E. S., Bean, M. G., & Stone, J. (2015). Confrontation and beyond: Examining a stigmatized target's use of a prejudice reduction strategy. *Social and Personality Psychology Compass, 9*(2), 100-114. https://doi.org/10.1111/spc3.12153.

Fuesting, M. A., Diekman, A. B., Boucher, K. L., Murphy, M. C., Manson, D. L., & Safer, B. L. (2019). Growing STEM: Perceived faculty mindset as an indicator of communal affordances in STEM. *Journal of Personality and Social Psychology, 117*(2), 260. https://psycnet.apa.org/doi/10.1037/pspa0000154.

Gervais, S. J., & Hillard, A. L. (2014). Confronting sexism as persuasion: Effects of a confrontation's recipient, source, message, and context. *Journal of Social Issues, 70*(4), 653-667. https://doi.org/10.1111/josi.12084.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700. https://psycnet.apa.org/doi/10.1037/a0026659.

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

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, Second Edition* (2nd ed.). The Guilford Press.

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures, 11*(1), 1-30. https://doi.org/10.1080/19312458.2016.1271116.

Hennes, E. P., Pietri, E. S., Moss-Racusin, C. A., Mason, K. A., Dovidio, J. F., Brescoll, V. L., Bailey, A.H., & Handelsman, J. (2018). Increasing the perceived malleability of gender bias using a modified video intervention for diversity in STEM (VIDS). Group Processes & Intergroup Relations, 21(5), 788-809. https://doi.org/10.1177%2F1368430218755923.

Hennessey, E. (2018). Confronting Sexism in Science, Technology, Engineering, and Math (STEM): What Are the Consequences? *Theses and Dissertations (Comprehensive)*. https://scholars.wlu.ca/etd/2054

Jones, E. E., & Nisbett, R. E. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, NJ: General Learning Press.

Kaiser, C. R., & Miller, C. T. (2001). Stop complaining! The social costs of making attributions to discrimination. *Personality and Social Psychology Bulletin*, *27*(2), 254-263. https://doi.org/10.1177%2F0146167201272010.

Karmali, F., Kawakami, K., & Page-Gould, E. (2017). He said what? Physiological and cognitive responses to imagining and witnessing outgroup racism. *Journal of Experimental Psychology: General*, *146*(8), 1073. https://psycnet.apa.org/doi/10.1037/xge0000304.

Kawakami, K., Dunn, E., Karmali, F., & Dovidio, J. F. (2009). Mispredicting affective and

behavioral responses to racism. *Science*, *323*(5911), 276-278.

https://doi.org/10.1126/science.1164951.

Kowalski, R. M., & Erickson, J. R. (1997). Complaining: What's all the fuss about?. In R. M.

Kowalski (Ed.), *Aversive Interpersonal Behaviors* (pp. 215–233). Springer US.

https://doi.org/ 10.1007/978-1-4757-9354-3_5

Kroeper, K. M. (2020). *Expecting prejudice confrontation to backfire: Prejudice norms and*

misalignment between forecaster expectations and experiencer realities (Doctoral

dissertation, Indiana University).

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.

LaCosse, J., Murphy, M. C., Garcia, J. A., & Zirkel, S. (2021). The role of STEM professors'

mindset beliefs on students' anticipated psychological experiences and course interest.

Journal of Educational Psychology, *113*(5), 949–971.

https://doi.org/10.1037/edu0000620

Lassetter, B., & Neel, R. (2019). Malleable liberals and fixed conservatives? Political

orientation shapes perceived ability to change. *Journal of Experimental Social*

Psychology, *82*, 141-151. https://doi.org/10.1016/j.jesp.2019.01.002.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the Mediation,

Confounding and Suppression Effect. *Prevention Science*, *1*(4), 173–181.

https://doi.org/10.1023/A:1026595011371

Malle, B. F. (20061030). The actor-observer asymmetry in attribution: A (surprising) meta-

analysis. *Psychological Bulletin*, *132*(6), 895. https://doi.org/10.1037/0033-

2909.132.6.895

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.

Muenks, K., Canning, E. A., LaCosse, J., Green, D. J., Zirkel, S., Garcia, J. A., & Murphy,

M. C. (2020). Does my professor think my ability can change? Students' perceptions

of their STEM professors' mindset beliefs predict their psychological vulnerability,

engagement, and performance in class. *Journal of Experimental Psychology: General,*

149(11), 2119–2144. https://doi.org/10.1037/xge0000763

Murphy, M. C., & Reeves, S. L. (2019). Personal and organizational mindsets at work.

Research in Organizational Behavior, 39, 100121.

https://doi.org/10.1016/j.riob.2020.100121.

Neel, R., & Lassetter, B. (2015). Growing fixed with age: Lay theories of malleability are

target age-specific. *Personality and Social Psychology Bulletin*, *41*(11), 1505-1522.

https://doi.org/10.1177%2F0146167215600529.

Neel, R., & Shapiro, J. R. (2012). Is racial bias malleable? Whites' lay theories of racial bias

predict divergent strategies for interracial interactions. *Journal of Personality and*

Social Psychology, *103*(1), 101. https://psycnet.apa.org/doi/10.1037/a0028237.

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.

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

Pettigrew, T. F., & Tropp, L. (2000). Does intergroup contact reduce racial and ethnic

prejudice throughout the world. *Reducing prejudice and discrimination. Mahwah, NJ: Lawrence Erlbaum.*

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

Rattan, A., & Dweck, C. S. (2010). Who confronts prejudice? The role of implicit theories in the motivation to confront prejudice. *Psychological Science*, *21*(7), 952-959. https://doi.org/10.1177%2F0956797610374740

Rattan, A., & Dweck, C. S. (2018). What happens after prejudice is confronted in the workplace? How mindsets affect minorities' and women's outlook on future social relations. *Journal of Applied Psychology, 103*(6), 676. https://psycnet.apa.org/doi/10.1037/apl0000287.

Rattan, A., Kroeper, K. M., Arnett, R., Brown, X., & Murphy, M. C. (2021, July 5). Mindsets and Prejudice Confrontation: Boundaries and Benefits of a Growth Mindset. Retrieved from osf.io/b7uxc

Rattan, A., & Ozgumus, E. (2019). Embedding mindsets in context: Theoretical considerations and opportunities for studying fixed-growth lay theories in the workplace. *Research in Organizational Behavior*, *39*, 100-127. https://doi.org/10.1016/j.riob.2020.100127.

Rattan, A., Savani, K., Komarraju, M., Morrison, M. M., Boggs, C., & Ambady, N. (2018). Meta-lay theories of scientific potential drive underrepresented students' sense of belonging to science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, *115*(1), 54. https://psycnet.apa.org/doi/10.1037/pspi0000130.

Rattan, A. (2019). How lay theories (or mindsets) shape the confrontation of prejudice.

In *Confronting Prejudice and Discrimination* (pp. 121-140). Academic Press.

Schultz, J. R., & Maddox, K. B. (2013). Shooting the messenger to spite the message? Exploring reactions to claims of racial bias. *Personality and Social Psychology Bulletin*, *39*(3), 346-358. https://doi.org/10.1177%2F0146167212475223.

J. Nicole, S., & Stewart, R. E. (2004). Confronting perpetrators of prejudice: The inhibitory effects of social costs. *Psychology of Women Quarterly*, *28*(3), 215-223. https://doi.org/10.1111%2Fj.1471-6402.2004.00138.x.

Simon, S., Shaffer, E., Neel, R., & Shapiro, J. (2019). Exploring Blacks' perceptions of Whites' racial prejudice as a function of intergroup behavior and motivational mindsets. *Social Psychological and Personality Science*, *10*(5), 575-585. https://doi.org/10.1177%2F1948550618778584.

Slepian, M. L., Young, S. G., Rutchick, A. M., & Ambady, N. (2013). Quality of professional players' poker hands is perceived accurately from arm motions. *Psychological Science*, *24*(11), 2335-2338. https://doi.org/10.1177%2F0956797613487384.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845–851. https://doi.org/10.1037/0022-3514.89.6.845

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/10.1016/j.jesp.2010.12.016.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). *People as flexible interpreters:*

Evidence and issues from spontaneous trait inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology, Vol. 28* (p. 211–279). Academic Press. https://doi.org/10.1016/S0065-2601(08)60239-7.

Uleman, J. S., Newman, L., & Winter, L. (1992). Can personality traits be inferred automatically? Spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*(1), 77–90. https://doi.org/10.1016/1053-8100(92)90049-G

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.